The Library of Scott Alexandria

Scott Alexander

The Library of Scott Alexandria

This is a collection of essays by Scott Alexander, collected by Rob Bensinger.

The ebook was assembled by Nino Annighöfer.

I. Rationality and Rationalization

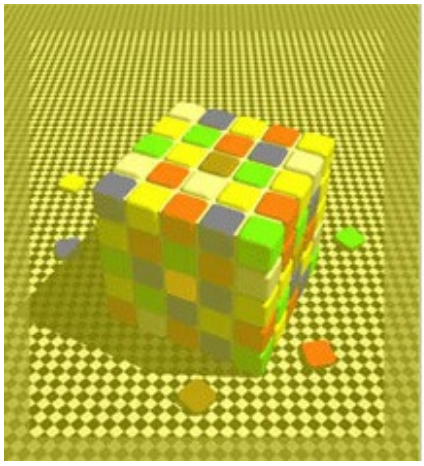
Blue- and Yellow-Tinted Choices

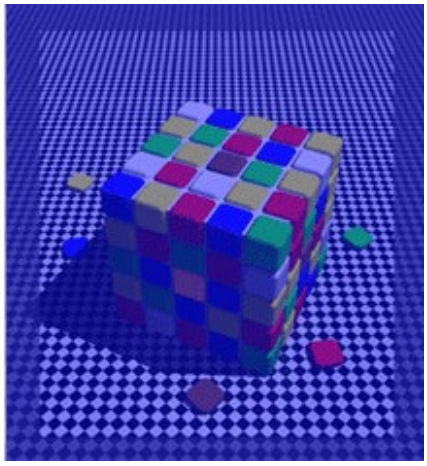
> A man comes to the rabbi and complains about his life: "I have almost no money, my wife is a shrew, and we live in a small apartment with seven unruly kids. It's messy, it's noisy, it's smelly, and I don't want to live."
> The rabbi says, "Buy a goat."
> "What? I just told you there's hardly room for nine people, and it's messy as it is!"
> "Look, you came for advice, so I'm giving you advice. Buy a goat and come back in a month."
> In a month the man comes back and he is even more depressed: "It's gotten worse! The filthy goat breaks everything, and it stinks and makes more noise than my wife and seven kids! What should I do?"
> The rabbi says, "Sell the goat."
> A few days later the man returns to the rabbi, beaming with happiness: "Life is wonderful! We enjoy every minute of it now that there's no goat - only the nine of us. The kids are well-behaved, the wife is agreeable - and we even have some money!"

— *traditional Jewish joke*

Related to: Anchoring and Adjustment

Biases are "cognitive illusions" that work on the same principle as optical illusions, and a knowledge of the latter can be profitably applied to the former. Take, for example, these two cubes (source: Lotto Lab, via Boing Boing):





The “blue” tiles on the top face of the left cube are the same color as the “yellow” tiles on the top face of the right cube; if you’re skeptical you can prove it with the eyedropper tool in Photoshop (in which both shades come out a rather ugly gray).

The illusion works because visual perception is relative. Outdoor light on a sunny day can be ten thousand times greater than a fluorescently lit indoor room. As one psychology book put it: for a student reading this book outside, the black print will be objectively lighter than the white space will be for a student reading the book inside. Nevertheless, both students will perceive the white space as subjectively white and the black space as subjectively black, because the visual system returns to consciousness information about relative rather than absolute lightness. In the two cubes, the visual system takes the yellow or blue tint as a given and outputs to consciousness the colors of each pixel compared to that background.

So this optical illusion occurs when the brain judges quantities relative to their surroundings rather than based on some objective standard. What’s the corresponding cognitive illusion?

In *Predictably Irrational* (relatively recommended, even though the latter chapters sort of fail to live up to the ones mentioned here) Dan Ariely asks his students to evaluate (appropriately) three subscription plans to the Economist:

Ariely asked his subjects which plan they'd buy if they needed an Economist subscription. 84% wanted the combo plan, 16% wanted the web only plan, and no one wanted the print only plan. After all, the print plan cost exactly the same as the print + web plan, but the print + web plan was obviously better. Which raises the question: why even include a print-only plan? Isn't it something of a waste of space?

Actually, including the print-only plan turns out to be a very good business move for the Economist. Ariely removed the print-only plan from the choices. Now the options looked like this.

There shouldn't be any difference. After all, he'd only removed the plan no one chose, the plan no sane person would choose.

This time, 68% of students chose the web only plan and 32% the combo plan. That's a 52% shift in preferences between the exact same options.

The rational way to make the decision is to compare the value of a print subscription to the Economist (as measured by the opportunity cost of that money) to the difference in cost between the web and combo subscriptions. But this would return the same answer in both of the above cases, so the students weren't doing it that way.

What it looks like the students were doing was perceiving relative value in the same way the eye perceives relative color. The ugly gray of the cube appeared blue when it was next to something yellow, and yellow when it was next to something blue. In the same way, the $125 cost of the combo subscription looks like good value next to a worse deal, and bad value next to a better deal.

When the $125 combo subscription was placed next to a $125 plan with fewer features (print only instead of print plus web) it looked like a very good deal – the equivalent of placing an ugly gray square next to something yellow to make it look blue. Take away the yellow, or the artificially bad deal, and it doesn't look nearly as attractive.

This is getting deep into Dark Arts territory, and according to Predictably Irrational, the opportunity to use these powers for evil has not gone unexploited. Retailers will deliberately include in their selection a super deluxe luxury model much fancier and more expensive than they expect anyone to ever want. The theory is that consumers are balancing a natural hedonism that tells them to get the best model possible against a commitment to financial prudence. So most consumers, however much they like

television, will have enough good sense to avoid buying a $2000 TV. But if the retailer carries a $4000 super-TV, the $2000 TV suddenly doesn't look quite so bad.

The obvious next question is "How do I use this knowledge to trick hot girls or guys into going out with me?" Dan Ariely decided to run some experiments on his undergraduate class. He took photographs of sixty students, then asked other students to rate their attractiveness. Next, he grouped the photos into pairs of equally attractive students. And next, he went to Photoshop and made a slightly less attractive version of each student: a blemish here, an asymmetry there.

Finally, he went around campus, finding students and showing them three photographs and asking which person the student would like to go on a date with. Two of the photographs were from one pair of photos ranked equally attractive. The third was a version of one of the two, altered to make it less attractive. So, for example, he might have two people, Alice and Brenda, who had been ranked equally attractive, plus a Photoshopped ugly version of Brenda.

The students overwhelmingly (75%) chose the person with the ugly double (Brenda in the example above), even though the two non-Photoshopped faces were equally attractive. Ariely then went so far as to recommend in his book that for best effect, you should go to bars and clubs with a wingman who is similar to you but less attractive. Going with a random ugly person would accomplish nothing, but going with someone similar to but less attractive than you would put you into a reference class and then bump you up to the top of the reference class, just like in the previous face experiment.

Ariely puts these studies in a separate chapter from his studies on anchoring and adjustment (which are also very good) but it all seems like the same process to me: being more interested in the

difference between two values than in the absolute magnitude of them. All that makes anchoring and adjustment so interesting is that the two values have nothing in common with one another.

This process also has applications to happiness set points, status seeking, morality, dieting, larger-scale purchasing behavior, and akrasia which deserve a separate post

What's In A Name?

> ***Marge:*** *You changed your name without consulting me?*
> ***Homer:*** *That's the way Max Power is, Marge. Decisive.*
> —The Simpsons

In honor of Will Powers and his theories about self-control, today I would like to talk about my favorite bias ever, the name letter effect. The name letter effect doesn't cause global existential risk or stock market crashes, and it's pretty far down on the list of things to compensate for. But it's a good example of just how insidious biases can be and of the egoism that permeates every level of the mind.

The name letter effect is your subconscious preference for things that sound like your own name. This might be expected to mostly apply to small choices like product brand names, but it's been observed in choices of spouse, city of residence, and even career. Some evidence comes from Pelham et al's Why Susie Sells Seashells By The Seashore:

The paper's first few studies investigate the relationship between a person's name and where they live. People named Phil were found more frequently than usual in Philadelphia, people named Jack in Jacksonville, people named George in Georgia, and so on with $p < .001$. To eliminate the possibility of the familiarity effect causing parents to subconsciously name their children after their place of residence, further studies were done with surnames and with people who moved later in life, both with the same results. The results held across US and Canadian city names as well as US state names, and were significant both for first name and surname.

In case that wasn't implausible enough, the researchers also looked at association between birth date and city of residence: that is, were people born on 2/02 more likely to live in the town of Two Harbors, and 3/03 babies more likely to live in Three Forks? With p = .003, yes, they are.

The researchers then moved on to career choices. They combed the records of the American Dental Association and the American Bar association looking for people named either Dennis, Denice, Dena, Denver, et cetera, or Lawrence, Larry, Laura, Lauren, et cetera. That is: were there more dentists named Dennis and lawyers named Lawrence than vice versa? Of the various statistical analyses they performed, most said yes, some at < .001 level. Other studies determined that there was a suspicious surplus of geologists named Geoffrey, and that hardware store owners were more likely to have names starting with 'H' compared to roofing store owners, who were more likely to have names starting with 'R'.

Some other miscellaneous findings: people are more likely to donate to Presidential candidates whose names begin with the same letter as their own, people are more likely to marry spouses whose names begin with the same letter as their own, that women are more likely to show name preference effects than men (but why?), and that batters with names beginning in 'K' are more likely than others to strike out (strikeouts being symbolized by a 'K' on the records).

If you have any doubts about the validity of the research, I urge you to read the linked paper. It's a great example of researchers who go above and beyond the call of duty to eliminate as many confounders as possible.

The name letter effect is a great addition to any list of psychological curiosities, but it does have some more solid

applications. I often use it as my first example when I'm introducing the idea of subconscious biases to people, because it's clear, surprising, and has major real-world effects. It also tends to shut up people who don't believe there are subconscious influences on decision-making, and who are always willing to find some excuse for why a supposed "bias" could actually be an example of legitimate decision-making.

And it introduces the concept of implicit egoism, the tendency to prefer something just because it's associated with you. It's one possible explanation for the endowment effect, and if it applies to *my* beliefs as strongly as to *my* personal details or *my* property, it's yet another mechanism by which opinions become calcified.

This is also an interesting window onto the complex and important world of self-esteem. Jones, Pelham et al suggest that the name preference effect is either involved in or a byproduct of some sort of self-esteem regulatory system. They find that name preferences are most common among high self-esteem people who have just experienced threats to their self-esteem, almost as if it is a reactive way of saying "No, you really are that great." I think an examination of how different biases interact with self-esteem would be a profitable direction for future research.

The Apologist and the Revolutionary

Rationalists complain that most people are too willing to make excuses for their positions, and too unwilling to abandon those positions for ones that better fit the evidence. And most people really *are* pretty bad at this. But certain stroke victims called anosognosiacs are much, much worse.

Anosognosia is the condition of not being aware of your own disabilities. To be clear, we're not talking minor disabilities here, the sort that only show up during a comprehensive clinical exam. We're talking paralysis or even blindness[1]. Things that should be pretty hard to miss.

Take the example of the woman discussed in Lishman's Organic Psychiatry. After a right-hemisphere stroke, she lost movement in her left arm but continuously denied it. When the doctor asked her to move her arm, and she observed it not moving, she claimed that it wasn't actually her arm, it was her daughter's. Why was her daughter's arm attached to her shoulder? The patient claimed her daughter had been there in the bed with her all week. Why was her wedding ring on her daughter's hand? The patient said her daughter had borrowed it. Where was the patient's arm? The patient "turned her head and searched in a bemused way over her left shoulder".

Why won't these patients admit they're paralyzed, and what are the implications for neurotypical humans? Dr. Vilayanur Ramachandran, leading neuroscientist and current holder of the world land-speed record for hypothesis generation, has a theory.

One immediately plausible hypothesis: the patient is unable to cope psychologically with the possibility of being paralyzed,

so he responds with denial. Plausible, but according to Dr. Ramachandran, wrong. He notes that patients with left-side strokes almost never suffer anosognosia, even though the left side controls the right half of the body in about the same way the right side controls the left half. There must be something special about the right hemisphere.

Another plausible hypothesis: the part of the brain responsible for thinking about the affected area was damaged in the stroke. Therefore, the patient has lost access to the area, so to speak. Dr. Ramachandran doesn't like this idea either. The lack of right-sided anosognosia in left-hemisphere stroke victims argues against it as well. But how can we disconfirm it?

Dr. Ramachandran performed an experiment[2] where he "paralyzed" an anosognosiac's good right arm. He placed it in a clever system of mirrors that caused a research assistant's arm to look as if it was attached to the patient's shoulder. Ramachandran told the patient to move his own right arm, and the false arm didn't move. What happened? The patient claimed he could see the arm moving - a classic anosognosiac response. This suggests that the anosognosia is not specifically a deficit of the brain's left-arm monitoring system, but rather some sort of failure of rationality.

Says Dr. Ramachandran:

> The reason anosognosia is so puzzling is that we have come to regard the 'intellect' as primarily propositional in character and one ordinarily expects propositional logic to be internally consistent. To listen to a patient deny ownership of her arm and yet, in the same breath, admit that it is attached to her shoulder is one of the most perplexing phenomena that one can encounter as a neurologist.

So what's Dr. Ramachandran's solution? He posits two different reasoning modules located in the two different hemispheres. The left brain tries to fit the data to the theory to preserve a coherent internal narrative and prevent a person from jumping back and forth between conclusions upon each new data point. It is primarily an apologist, there to explain why any experience is exactly what its own theory would have predicted. The right brain is the seat of the second virtue. When it's had enough of the left-brain's confabulating, it initiates a Kuhnian paradigm shift to a completely new narrative. Ramachandran describes it as "a left-wing revolutionary".

Normally these two systems work in balance. But if a stroke takes the revolutionary offline, the brain loses its ability to change its mind about anything significant. If your left arm was working before your stroke, the little voice that ought to tell you it might be time to reject the "left arm works fine" theory goes silent. The only one left is the poor apologist, who must tirelessly invent stranger and stranger excuses for why all the facts really fit the "left arm works fine" theory perfectly well.

It gets weirder. For some reason, squirting cold water into the left ear canal wakes up the revolutionary. Maybe the intense sensory input from an unexpected source makes the right hemisphere unusually aroused. Maybe distoring the balance sense causes the eyes to move rapidly, activating a latent system for inter-hemisphere co-ordination usually restricted to REM sleep[3]. In any case, a patient who has been denying paralysis for weeks or months will, upon having cold water placed in the ear, admit to paralysis, admit to having been paralyzed the past few weeks or months, and express bewilderment at having ever denied such an obvious fact. And

then the effect wears off, and the patient not only denies the paralysis but denies ever having admitted to it.

This divorce between the apologist and the revolutionary might also explain some of the odd behavior of split-brain patients. Consider the following experiment: a split-brain patient was shown two images, one in each visual field. The left hemisphere received the image of a chicken claw, and the right hemisphere received the image of a snowed-in house. The patient was asked verbally to describe what he saw, activating the left (more verbal) hemisphere. The patient said he saw a chicken claw, as expected. Then the patient was asked to point with his left hand (controlled by the right hemisphere) to a picture related to the scene. Among the pictures available were a shovel and a chicken. He pointed to the shovel. So far, no crazier than what we've come to expect from neuroscience.

Now the doctor verbally asked the patient to describe why he just pointed to the shovel. The patient verbally (left hemisphere!) answered that he saw a chicken claw, and of course shovels are necessary to clean out chicken sheds, so he pointed to the shovel to indicate chickens. The apologist in the left-brain is helpless to do anything besides explain why the data fits its own theory, and its own theory is that whatever happened had something to do with chickens, dammit!

The logical follow-up experiment would be to ask the right hemisphere to explain the left hemisphere's actions. Unfortunately, the right hemisphere is either non-linguistic or as close as to make no difference. Whatever its thoughts, it's keeping them to itself.

…you know, my mouth is *still* agape at that whole cold-water-in-the-ear trick. I have this fantasy of gathering all the leading

creationists together and squirting ice cold water in each of their left ears. All of a sudden, one and all, they admit their mistakes, and express bafflement at ever having believed such nonsense. And then ten minutes later the effect wears off, and they're all back to talking about irreducible complexity or whatever. I don't mind. I've already run off to upload the video to YouTube.

This is surely so great an exaggeration of Dr. Ramachandran's theory as to be a parody of it. And in any case I don't know how much to believe all this about different reasoning modules, or how closely the intuitive understanding of it I take from his paper matches the way a neuroscientist would think of it. Are the apologist and the revolutionary active in normal thought? Do anosognosiacs demonstrate the same pathological inability to change their mind on issues other than their disabilities? What of the argument that confabulation is a rather common failure mode of the brain, shared by some conditions that have little to do with right-hemisphere failure? Why does the effect of the cold water wear off so quickly? I've yet to see any really satisfying answers to any of these questions.

But whether Ramachandran is right or wrong, I give him enormous credit for doing serious research into the neural correlates of human rationality. I can think of few other fields that offer so many potential benefits.

Footnotes

1: See Anton-Babinski syndrome

2: See Ramachandran's "The Evolutionary Biology of Self-Deception", the link from "posits two different reasoning modules" in this article.

3: For Ramachandran's thoughts on REM, again see "The Evolutionary Biology of Self Deception"

Historical realism

As I mentioned in my last entry, I've been watching Babylon 5 lately. It's not a perfect show, but it has one big advantage: it's consistent and believable.

Contrast this with Doctor Who. Doctor Who is fun to watch, but if you think about it for more than two seconds you notice it's full of plot holes and contradictions. Things that cause time travel paradoxes that threaten to destroy the universe one episode go without a hitch the next. And the TARDIS, the sonic screwdriver, and the Doctor's biology gain completely different powers no one's ever alluded to depending on the situation. The aliens are hysterically unlikely, often without motives or believable science, the characters will do any old insane thing when it makes the plot slightly more interesting, and everything has either a self-destruct button or an easily findable secret weakness that it takes no efforts to defend against.

But I guess I'm not complaining. If the show was believable, the Doctor would have gotten killed the first time he decided to take on a massive superadvanced alien invasion force by walking right up to them openly with no weapons and no plan. And then they would have had to cancel the show, and then I would lose my chance to look at the pretty actress who plays Amy Pond.

So Doctor Who is not a complete loss. But then there are some shows that go completely beyond the pale of enjoyability, until they become nothing more than overwritten collections of tropes impossible to watch without groaning.

I think the worst offender here is the History Channel and all their programs on the so-called “World War II”.

Let’s start with the bad guys. Battalions of stormtroopers dressed in all black, check. Secret police, check. Determination to brutally kill everyone who doesn’t look like them, check. Leader with a tiny villain mustache and a tendency to go into apopleptic rage when he doesn’t get his way, check. All this from a country that was ordinary, believable, and dare I say it sometimes even *sympathetic* in previous seasons.

I wouldn’t even mind the lack of originality if they weren’t so heavy-handed about it. Apparently we’re supposed to believe that in the middle of the war the Germans attacked their allies the Russians, starting an unwinnable conflict on two fronts, just to show how sneaky and untrustworthy they could be? And that they diverted all their resources to use in making ever bigger and scarier death camps, even in the middle of a huge war? Real people *just aren’t that evil*. And that’s not even counting the part where as soon as the plot requires it, they instantly forget about all the racism nonsense and become best buddies with the definitely non-Aryan Japanese.

Not that the good guys are much better. Their leader, Churchill, appeared in a grand total of one episode before, where he was a bumbling general who suffered an embarrassing defeat to the Ottomans of all people in the Battle of Gallipoli. Now, all of a sudden, he’s not only Prime Minister, he’s not only a brilliant military commander, he’s not only the greatest orator of the twentieth century who can convince the British to keep going against all odds, he’s *also* a natural wit who is able to pull out hilarious one-liners practically on demand. I know he’s supposed to be the hero,

but it's not realistic unless you keep the guy at least vaguely human.

So it's pretty standard "shining amazing good guys who can do no wrong" versus "evil legions of darkness bent on torture and genocide" stuff, totally ignoring the nuances and realities of politics. The actual strategy of the war is barely any better. Just to give one example, in the Battle of the Bulge, a vastly larger force of Germans surround a small Allied battalion and demand they surrender or be killed. The Allied general sends back a single-word reply: "Nuts!". The Germans attack, and, miraculously, the tiny Allied force holds them off long enough for reinforcements to arrive and turn the tide of battle. Whoever wrote this episode obviously had never been within a thousand miles of an actual military.

Probably the worst part was the ending. The British/German story arc gets boring, so they tie it up quickly, have the villain kill himself (on Walpurgisnacht of all days, not exactly subtle) and then totally switch gears to a battle between the Americans and the Japanese in the Pacific. Pretty much the same dichotomy - the Japanese kill, torture, perform medical experiments on prisoners, and frickin' play football with the heads of murdered children, and the Americans are led by a kindly old man in a wheelchair.

Anyway, they spend the whole season building up how the Japanese home islands are a fortress, and the Japanese will never surrender, and there's no way to take the Japanese home islands because they're invincible…and then they realize they totally can't have the Americans take the Japanese home islands so they have no way to wrap up the season.

So they invent a completely implausible superweapon that they've *never* mentioned until now. Apparently the Americans

got some scientists together to invent it, only we never heard anything about it because it was “classified”. In two years, the scientists manage to invent a weapon a thousand times more powerful than anything anyone’s ever seen before - drawing from, of course, ancient mystical texts. Then they use the superweapon, blow up several Japanese cities easily, and the Japanese surrender. Convenient, isn’t it?

…and then, in the entire rest of the show, over five or six different big wars, they never use the superweapon again. Seriously. They have this whole thing about a war in Vietnam that lasts decades and kills tens of thousands of people, and they never wonder if maybe they should consider using *the frickin’ unstoppable mystical superweapon that they won the last war with*. At this point, you’re starting to wonder if any of the show’s writers have even *watched* the episodes the other writers made.

I’m not even going to get into the whole subplot about breaking a secret code (cleverly named “Enigma”, because the writers couldn’t spend more than two seconds thinking up a name for an enigmatic code), the giant superintelligent computer called Colossus (despite this being years before the transistor was even *invented*), the Soviet strongman whose name means “Man of Steel” in Russian (seriously, between calling the strongman “Man of Steel” and the Frenchman “de Gaulle”, whoever came up with the names for this thing ought to be shot).

So yeah. Stay away from the History Channel. Unlike most of the other networks, they don’t even *try* to make their stuff believable.

Simultaneously Right and Wrong

Related to: Belief in Belief, Convenient Overconfidence

"You've no idea of what a poor opinion I have of myself, and how little I deserve it."

— W.S. Gilbert

In 1978, Steven Berglas and Edward Jones performed a study on voluntary use of performance inhibiting drugs. They asked subjects to solve certain problems. The control group received simple problems, the experimental group impossible problems. The researchers then told all subjects they'd solved the problems successfully, leaving the controls confident in their own abilities and the experimental group privately aware they'd just made a very lucky guess.

Then they offered the subjects a choice of two drugs to test. One drug supposedly enhanced performance, the other supposedly handicapped it.

There's a cut here in case you want to predict what happened.

Males in the lucky guesser group chose the performance-inhibiting drug significantly more than those in the control group[1].

The researchers conjectured that the lucky guessers felt good about acing the first test. They anticipated failing the second whether they chose the helpful or the harmful drug. So they chose the harmful drug as an excuse: "Oh, I would have passed the test, only the drug was making me stupid." As the study points out, this is a win-win situation: if they fail, the drug excuses their failure, and if they succeed it's doubly impressive that they passed even with a handicap.

Since this study, psychologists have applied self-handicapping to interpret common phenomena like procrastination, hypochondria, and minority underachievement; they've also discovered factors that increase or decrease self-handicapping tendencies (for example, you can increase it by making the test more relevant to your self-esteem: just say "this is a proven test of intellectual ability in general").

But some especially interesting studies investigated the effect of privacy on self-handicapping. For example, Hobden conducted an experiment similar to Berglas and Jones', albeit with tapes of performance-enhancing or -handicapping music instead of drugs. The twist: half the subjects' choice of tapes and test scores were public, and half the subjects believed their choice of tapes and test scores were anonymous. What happens when no one but the subject himself will ever know his test score? He self-handicaps just as often as everyone else. And it seems to *work*. The same set of studies showed that subjects who self-handicap on a test are less likely to attribute their failure on the test to their own incompetence.

In order to handicap, subjects must have an inaccurate assessment of their own abilities. Otherwise, there's no self-esteem to protect. If I believe my IQ is 80, and I get 80 on an IQ test, I have no incentive to make excuses to myself, or to try to explain away the results. The only time I would want to explain away the results as based on some external factor was if I'd been going around thinking my real IQ was 100.

But subjects also must have an accurate assessment of their own abilities. Subjects who take an easy pre-test and expect an easy test do not self-handicap. Only subjects who understand their low chances of success can think "I will probably fail this test, so I will need an excuse[2].

If this sounds familiar, it's because it's another form of the dragon problem from Belief in Belief. The believer says there is a dragon in his garage, but expects all attempts to detect the dragon's presence to fail. Eliezer writes: "The claimant must have an accurate model of the situation somewhere in his mind, because he can anticipate, in advance, exactly which experimental results he'll need to excuse."

Should we say that the subject believes he will get an 80, but believes in believing that he will get a 100? This doesn't quite capture the spirit of the situation. Classic belief in belief seems to involve value judgments and complex belief systems, but self-handicapping seems more like simple overconfidence bias[3]. Is there any other evidence that overconfidence has a belief-in-belief aspect to it?

Last November, Robin described a study where subjects were less overconfident if asked to predict their performance on tasks they will actually be expected to complete. He ended by noting that "It is almost as if we at some level realize that our overconfidence is unrealistic."

Belief in belief in religious faith and self-confidence seem to be two areas in which we can be simultaneously right and wrong: expressing a biased position on a superficial level while holding an accurate position on a deeper level. The specifics are different in each case, but perhaps the same general mechanism may underlie both. How many other biases use this same mechanism?

Footnotes

1: In most studies on this effect, it's most commonly observed among males. The reasons are too complicated and controversial to be discussed in this post, but are left as an

exercise for the reader with a background in evolutionary psychology.

2: Compare the ideal Bayesian, for whom expected future expectation is always the same as the current expectation, and investors in an ideal stock market, who must always expect a stock's price tomorrow to be on average the same as its price today - to this poor creature, who accurately predicts that he will lower his estimate of his intelligence after taking the test, but who doesn't use that prediction to change his pre-test estimates.

3: I have seen "overconfidence bias" used in two different ways: to mean poor calibration on guesses (ie predictions made with 99% certainty that are only right 70% of the time) and to mean the tendency to overestimate one's own good qualities and chance of success. I am using the latter definition here to remain consistent with the common usage on Overcoming Bias; other people may call this same error "optimism bias".

You May Already Be A Sinner

Followup to: Simultaneously Right and Wrong

Related to: Augustine's Paradox of Optimal Repentance

> *"When they inquire into predestination, they are penetrating the sacred precincts of divine wisdom. If anyone with carefree assurance breaks into this place, he will not succeed in satisfying his curiosity and he will enter a labyrinth from which he can find no exit."*
>
> — John Calvin

John Calvin preached the doctrine of predestination: that God irreversibly decreed each man's eternal fate at the moment of Creation. Calvinists separate mankind into two groups: the elect, whom God predestined for Heaven, and the reprobate, whom God predestined for eternal punishment in Hell.

If you had the bad luck to be born a sinner, there is nothing you can do about it. You are too corrupted by original sin to even have the slightest urge to seek out the true faith. Conversely, if you were born one of the elect, you've got it pretty good; no matter what your actions on Earth, it is impossible for God to revoke your birthright to eternal bliss.

However, it is believed that the elect always live pious, virtuous lives full of faith and hard work. Also, the reprobate always commit heinous sins like greed and sloth and commenting on anti-theist blogs. This isn't what causes God to damn them. It's just what happens to them after they've been damned: their soul has no connection with God and so it tends in the opposite direction.

Consider two Calvinists, Aaron and Zachary, both interested only in maximizing his own happiness. Aaron thinks to himself "Whether or not I go to Heaven has already been decided, regardless of my actions on Earth. Therefore, I might as well try to have as much fun as possible, knowing it won't effect the afterlife either way." He spends his days in sex, debauchery, and anti-theist blog comments.

Zachary sees Aaron and thinks "That sinful man is thus proven one of the reprobate, and damned to Hell. I will avoid his fate by living a pious life." Zachary becomes a great minister, famous for his virtue, and when he dies his entire congregation concludes he must have been one of the elect.

Before the cut: If you were a Calvinist, which path would you take?

Amos Tversky, Stanford psychology professor by day, bias-fighting superhero by night, thinks you should live a life of sin. He bases his analysis of the issue on the famous maxim that correlation is not causation. Your virtue during life is correlated to your eternal reward, but only because they're both correlated to a hidden third variable, your status as one of the elect, which causes both.

Just to make that more concrete: people who own more cars live longer. Why? Rich people buy more cars, and rich people have higher life expectancies. Both cars and long life are caused by a hidden third variable, wealth. Trying to increase your chances of getting into Heaven by being virtuous is as futile as trying to increase your life expectancy by buying another car.

Some people would stop there, but not Amos Tversky, bias-fighting superhero. He and George Quattrone conducted a study that both illuminated a flaw in human reasoning about

causation and demonstrated yet another way people can be simultaneously right and wrong.

Subjects came in thinking it was a study on cardiovascular health. First, experimenters tested their pain tolerance by making them stick their hands in a bucket of freezing water until they couldn't bear it any longer. However long they kept it there was their baseline pain tolerance score.

Then experimenters described two supposed types of human heart: Type I hearts, which work poorly and are prone to heart attack and will kill you at a young age, and Type II hearts, which work well and will bless you with a long life. You can tell a Type I heart from a Type II heart because…and here the subjects split into two groups. Group A learned that people with Type II hearts, the good hearts, had higher pain tolerance after exercise. Group B learned that Type II hearts had lower pain tolerance after exercise.

Then the subjects exercised for a while and stuck their hands in the bucket of ice water again. Sure enough, the subjects who thought increased pain tolerance meant a healthier heart kept their hands in longer. And then when the researchers went and asked them, they said they must have a Type II heart because the ice water test went so well!

The subjects seem to have believed *on some level* that keeping their hand in the water longer could give them a different kind of heart. Dr. Tversky declared that people have a cognitive blind spot to "hidden variable" causation, and this explains the Calvinists who made such an effort to live virtuously.

But this study is also interesting as an example of self-deception. One level of the mind made the (irrational) choice to leave the hand in the ice water longer. Another level of the mind that wasn't consciously aware of this choice interpreted

it as evidence for the Type II heart. There are two cognitive flaws here: the subject's choice to try harder on the ice water test, and his lack of realization that he'd done so.

I don't know of any literature explicitly connecting this study to self-handicapping, but the surface similarities are striking. In both, a person takes an action intended to protect his self-image that will work if and only if he doesn't realize this intent. In both, the action is apparently successful, self-image is protected, and the conscious mind remains unaware of the true motives.

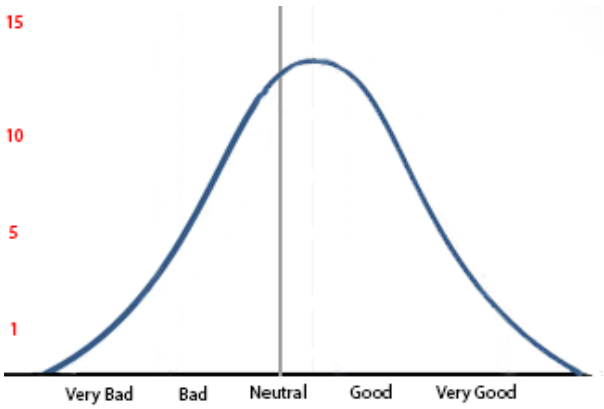
Despite all this, and with all due respect to Dr. Tversky I think he might be wrong about the whole predestination issue. If I were a Calvinist, I'd live a life of sin if and only if I would two-box on Newcomb's Problem.

Beware the Man of One Study

Aquinas famously said: beware the man of one book. I would add: beware the man of one study.

For example, take medical research. Suppose a certain drug is weakly effective against a certain disease. After a few years, a bunch of different research groups have gotten their hands on it and done all sorts of different studies. In the best case scenario the average study will find the true result – that it's weakly effective.

But there will also be random noise caused by inevitable variation and by some of the experiments being better quality than others. In the end, we might expect something looking kind of like a bell curve. The peak will be at "weakly effective", but there will be a few studies to either side. Something like this:



We see that the peak of the curve is somewhere to the right of neutral – ie weakly effective – and that there are about 15 studies that find this correct result.

But there are also about 5 studies that find that the drug is very good, and 5 studies missing the sign entirely and finding that the drug is actively bad. There's even 1 study finding that the drug is very bad, maybe seriously dangerous.

This is before we get into fraud or statistical malpractice. I'm saying this is what's going to happen just by normal variation in experimental design. As we increase experimental rigor, the bell

curve might get squashed horizontally, but there will still be a bell curve.

In practice it's worse than this, because this is assuming everyone is investigating exactly the same question.

Suppose that the graph is titled "Effectiveness Of This Drug In Treating Bipolar Disorder".

But maybe the drug is more effective in bipolar i than in bipolar ii (Depakote, for example)

Or maybe the drug is very effective against bipolar mania, but much less effective against bipolar depression (Depakote again).

Or maybe the drug is a good acute antimanic agent, but very poor at maintenance treatment (let's stick with Depakote).

If you have a graph titled "Effectiveness Of Depakote In Treating Bipolar Disorder" plotting studies from "Very Bad" to "Very Good" – and you stick all the studies – maintenence, manic, depressive, bipolar i, bipolar ii – on the graph, then you're going to end running the gamut from "very bad" to "very good" even before you factor in noise and even before even before you factor in bias and poor experimental design.

So here's why you should beware the man of one study.

If you go to your better class of alternative medicine websites, they don't tell you "Studies are a logocentric phallocentric tool of Western medicine and the Big Pharma conspiracy."

They tell you "medical science has proved that this drug is terrible, but ignorant doctors are pushing it on you anyway. Look, here's a study by a reputable institution proving that the drug is not only ineffective, but harmful."

And the study will exist, and the authors will be prestigious scientists, and it will probably be about as rigorous and well-done as any other study.

And then a lot of people raised on the idea that some things have Evidence and other things have No Evidence think *holy s**t, they're right!*

On the other hand, your doctor isn't going to a sketchy alternative medicine website. She's examining the entire literature and extracting careful and well-informed conclusions from…

Haha, just kidding. She's going to a luncheon at a really nice restaurant sponsored by a pharmaceutical company, which assures her that they would *never* take advantage of such an opportunity to shill their drug, they just want to raise awareness of the latest study. And the latest study shows that their drug is great! Super great! And your doctor nods along, because the authors of the study are prestigious scientists, and it's about as rigorous and well-done as any other study.

But obviously the pharmaceutical company has selected one of the studies from the "very good" end of the bell curve.

And I called this "Beware The Man of One Study", but it's easy to see that in the little diagram there are like three or four studies showing that the drug is "very good", so if your doctor is a little skeptical, the pharmaceutical company can say "You are right to be skeptical, one study doesn't prove anything, but look – here's another group that finds the same thing, here's yet another group that finds the same thing, and here's a replication that confirms both of them."

And even though it looks like in our example the sketchy alternative medicine website only has one "very bad" study to go off of, they could easily supplement it with a bunch of merely "bad" studies. Or they could add all of those studies about slightly different things. Depakote is ineffective at treating bipolar depression. Depakote is ineffective at maintenance bipolar therapy. Depakote is ineffective at bipolar ii.

So just sum it up as "Smith et al 1987 found the drug ineffective, yet doctors continue to prescribe it anyway". Even if you hunt down the

original study (which no one does), Smith et al won't say specifically "Do remember that this study is only looking at bipolar maintenance, which is a different topic from bipolar acute antimanic treatment, and we're not saying anything about that." It will just be titled something like "Depakote fails to separate from placebo in six month trial of 91 patients" and trust that the responsible professionals reading it are well aware of the difference between acute and maintenance treatments (hahahahaha).

So it's not so much "beware the man of one study" as "beware the man of any number of studies less than a relatively complete and not-cherry-picked survey of the research".

II.

I think medical science is still pretty healthy, and that the consensus of doctors and researchers is more-or-less right on most controversial medical issues.

(it's the *uncontroversial* ones you have to worry about)

Politics doesn't have this protection.

Like, take the minimum wage question (please). We all know about the Krueger and Card study in New Jersey that found no evidence that high minimum wages hurt the economy. We probably also know the counterclaims that it was completely debunked as despicable dishonest statistical malpractice. Maybe some of us know Card and Krueger wrote a pretty convincing rebuttal of those claims. Or that a bunch of large and methodologically advanced studies have come out since then, some finding no effect like Dube, others finding strong effects like Rubinstein and Wither. These are just examples; there are at least dozens and probably hundreds of studies on both sides.

But we can solve this with meta-analyses and systemtic reviews, right?

Depends which one you want. Do you go with this meta-analysis of fourteen studies that shows that any presumed negative effect of high minimum wages is likely publication bias? With this meta-

analysis of sixty-four studies that finds the same thing and discovers no effect of minimum wage after correcting for the problem? Or how about this meta-analysis of fifty-five countries that does find effects in most of them? Maybe you prefer this systematic review of a hundred or so studies that finds strong and consistent effects?

Can we trust news sources, think tanks, econblogs, and other institutions to sum up the state of the evidence?

CNN claims that 85% of credible studies have shown the minimum wage causes job loss. But raisetheminimumwage.com declares that "two decades of rigorous economic research have found that raising the minimum wage does not result in job loss…researchers and businesses alike agree today that the weight of the evidence shows no reduction in employment resulting from minimum wage increases." Modeled Behavior says "the majority of the new minimum wage research supports the hypothesis that the minimum wage increases unemployment." The Center for Budget and Policy Priorities says "The common claim that raising the minimum wage reduces employment for low-wage workers is one of the most extensively studied issues in empirical economics. The weight of the evidence is that such impacts are small to none."

Okay, fine. What about economists? They seem like experts. What do they think?

Well, five hundred economists signed a letter to policy makers saying that the science of economics shows increasing the minimum wage would be a bad idea. That sounds like a promising consensus…

..except that six hundred economists signed a letter to policy makers saying that the science of economics shows increasing the minimum wage would be a *good* idea. (h/t Greg Mankiw)

Fine then. Let's do a formal survey of economists. Now what?

raisetheminimumwage.com, an unbiased source if ever there was one, confidently tells us that "indicative is a 2013 survey by the University of Chicago's Booth School of Business in which leading

economists agreed by a nearly 4 to 1 margin that the benefits of raising and indexing the minimum wage outweigh the costs."

But the Employment Policies Institute, which sounds like it's trying *way* too hard to sound like an unbiased source, tells us that "Over 73 percent of AEA labor economists believe that a significant increase will lead to employment losses and 68 percent think these employment losses fall disproportionately on the least skilled. Only 6 percent feel that minimum wage hikes are an efficient way to alleviate poverty."

So the whole thing is fiendishly complicated. But unless you look very very hard, you will never know that.

If you are a conservative, what you will find on the sites you trust will be something like this:

> Economic theory has always shown that minimum wage increases decrease employment, but the Left has never been willing to accept this basic fact. In 1992, they trumpeted a single study by Card and Krueger that purported to show no negative effects from a minimum wage increase. This study was immediately debunked and found to be based on statistical malpractice and "massaging the numbers". Since then, dozens of studies have come out confirming what we knew all along – that a high minimum wage is economic suicide. Systematic reviews and meta-analyses (Neumark 2006, Boockman 2010) consistently show that an overwhelming majority of the research agrees on this fact – as do 73% of economists. That's why five hundred top economists recently signed a letter urging policy makers not to buy into discredited liberal minimum wage theories. Instead of listening to starry-eyed liberal woo, listen to the empirical evidence and an overwhelming majority of economists and oppose a raise in the minimum wage.

And if you are a leftist, what you will find on the sites you trust will be something like this:

> People used to believe that the minimum wage decreased unemployment. But Card and Krueger's famous 1992 study exploded that conventional wisdom. Since then, the results have been replicated over fifty times, and further meta-analyses (Card and Krueger 1995, Dube 2010) have found no evidence of any effect. Leading economists agree by a 4 to 1 margin that the benefits of raising the minimum wage outweigh the costs, and that's why more than 600 of them have signed a petition telling the government to do exactly that. Instead of listening to conservative scare tactics based on long-debunked theories, listen to the empirical evidence and the overwhelming majority of economists and support a raise in the minimum wage.

Go ahead. Google the issue and see what stuff comes up. If it doesn't quite match what I said above, it's usually because they can't even muster *that* level of scholarship. Half the sites just cite Card and Krueger and call it a day!

These sites with their long lists of studies and experts are super convincing. And half of them are wrong.

At some point in their education, most smart people usually learn not to credit arguments from authority. If someone says "Believe me about the minimum wage because I seem like a trustworthy guy," most of them will have at least one neuron in their head that says "I should ask for some evidence". If they're *really* smart, they'll use the magic words "peer-reviewed experimental studies."

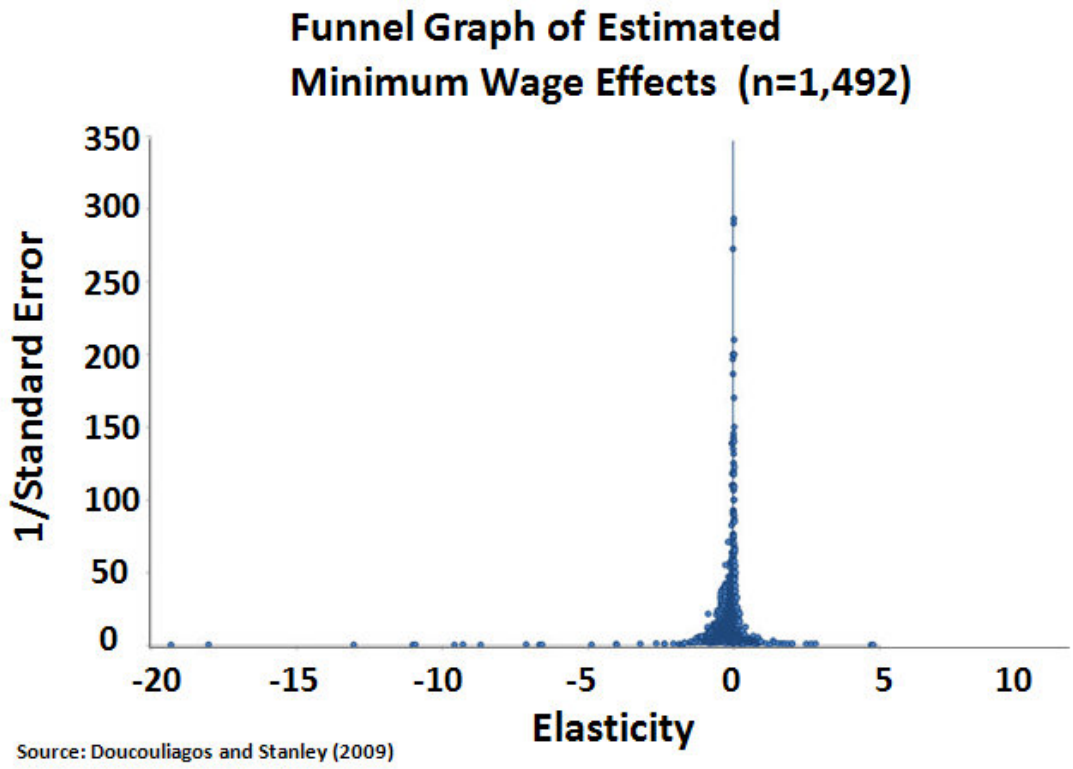
But I worry that most smart people have *not* learned that a list of dozens of studies, several meta-analyses, hundreds of experts, and expert surveys showing almost all academics support your thesis – can *still* be bullshit.

Which is too bad, because that's exactly what people who want to bamboozle an educated audience are going to use.

III.

I do not want to preach radical skepticism.

For example, on the minimum wage issue, I notice only one side has presented a funnel plot. A funnel plot is usually used to investigate publication bias, but it has another use as well – it's pretty much an exact presentation of the "bell curve" we talked about above.



This is more of a needle curve than a bell curve, but the point still stands. We see it's centered around 0, which means there's some evidence that's the real signal among all this noise. The bell skews more to left than to the right, which means more studies have found negative effects of the minimum wage than positive effects of the minimum wage. But since the bell curve is asymmetrical, we intepret that as *probably* publication bias. So all in all, I think there's at least some evidence that the liberals are right on this one.

Unless, of course, someone has realized that I've wised up to the studies and meta-analyses and and expert surveys, and figured out a way to hack *funnel plots*, which I am totally not ruling out.

(okay, I *kind of* want to preach radical skepticism)

Also, I should probably mention that it's much more complicated than one side being right, and that the minimum wage probably

works differently depending on what industry you're talking about, whether it's state wage or federal wage, whether it's a recession or a boom, whether we're talking about increasing from $5 to $6 or from $20 to $30, etc, etc, etc. There are eleven studies on that plot showing an effect even worse than -5, and very possibly they are all accurate for whatever subproblem they have chosen to study – much like the example with Depakote where it might an effective antimanic but a terrible antidepressant.

(radical skepticism actually sounds a lot better than figuring this all out).

IV.

But the question remains: what happens when (like in most cases) you don't have a funnel plot?

I don't have a good positive answer. I do have several good *negative* answers.

Decrease your confidence about most things if you're not sure that you've investigated every piece of evidence.

Do not trust websites which are obviously biased (eg Free Republic, Daily Kos, Dr. Oz) when they tell you they're going to give you "the state of the evidence" on a certain issue, even if the evidence seems very stately indeed. This goes double for any site that contains a list of "myths and facts about X", quadruple for any site that uses phrases like "ingroup member uses actual FACTS to DEMOLISH the outgroup's lies about Y", and octuple for RationalWiki.

Most important, even if someone gives you what seems like overwhelming evidence in favor of a certain point of view, don't trust it until you've done a simple Google search to see if the opposite side has equally overwhelming evidence.

Debunked and Well-Refuted

I.

As usual, I was insufficiently pessimistic.

I infer this from *The Federalist*'s article on campus rape:

> A new report on sexual assault released today by the U.S. Department of Justice (DOJ) officially puts to bed the bogus statistic that one in five women on college campuses are victims of sexual assault. In fact, non-students are 25 percent more likely to be victims of sexual assault than students, according to the data. And the real number of assault victims is several orders of magnitude lower than one-in-five.

The article compares the older Campus Sexual Assault Survey (which found 14-20% of women were raped since entering college) to the just-released National Crime Victmization Survey (which found that 0.6% of female college students are raped per year). They write "Instead of 1 in 5, the real number is 0.03 in 5."

So the first thing I will mock *The Federalist* for doing is directly comparing per year sexual assault rates to per college career sexual assault rates, whereas obviously these are very different things. You can't *quite* just divide the latter by four to get the former, but that's going to work a heck of a lot better than *not* doing it, so let's estimate the real discrepancy as more like 0.5% per year versus 5% per year.

But I can't get too mad at them yet, because that's still a pretty big discrepancy.

However, faced with this discrepancy a reasonable person might say "Hmm, we have two different studies that say two different things. I wonder what's going on here and which study we should believe?"

The Federalist staff said "Ha! There's an old study with findings we didn't like, but now there's a new study with different findings we *do* like. So the old study is debunked!"

II.

My last essay, Beware The Man Of One Study, noted that one thing partisans do to justify their bias is selectively acknowledge studies from only one side of a complicated literature.

The reason it was insufficiently pessimistic is that there are also people like the Federalist staff, who acknowledge the existence of opposing studies, but only with the adjective "debunked" in front of them. By "debunked" they usually mean one of two things:

1. Someone on my side published a study later that found something else
2. Someone on my side accused it of having methodological flaws

Since the Federalist has so amply demonstrated the first failure mode, let me say a little more about the second. Did you know that *anyone* with a keyboard can just *type up* any of the following things?

– "That study is a piece of garbage that's not worth the paper it's written on."
– "People in the know dismissed that study years ago."
– "Nobody in the field takes that study seriously."
– "That study uses methods that are laughable to anybody who

knows statistics."
– "All the other research that has come out since discredits that study."

They can say these things *whether they are true or not*. I'm kind of harping on this point, but it's because it's something *I* didn't realize until much later than I should have.

There are many "questions" that are pretty much settled – evolution, global warming, homeopathy. But taking these as representative closes your mind and gives you a skewed picture of academia. On many issues, academics are just as divided as anyone else, and their arguments can be just as acrimonious as anyone else's. The arguments usually take the form of one side publishing a study, the other side ripping the study apart and publishing their own study which they say is better, and the first side ripping the second study apart and arguing that their study was better all along.

Every study has flaws. No study has perfect methodology. If you like a study, you can say that it did the best it could on a difficult research area and has improved upon even-worse predecessor studies. If you don't like a study, you can say "LOOK AT THESE FLAWS THESE PEOPLE ARE IDIOTS THE CONCLUSION IS COMPLETELY INVALID". All you need to do is make enough isolated demands for rigor against anything you disagree with.

And so if the first level of confirmation bias is believing every study that supports your views, the second layer of confirmation bias is believing every supposed refutation that supports your views.

See for example this recent Xenosystems post about a Twitterer claiming *The Bell Curve* has been "well-refuted". There are definitely a lot of people who have written books,

articles, and papers arguing that *The Bell Curve* is wrong, often in very strong terms. There are also a lot of people who have written books, articles, and papers saying that the first set of books, articles, and papers are wrong and *The Bell Curve* is right, also in very strong terms. To say that the first set is a "refutation" or "debunking" is as basic a mistake as saying that the new rape study is a "refutation" or "debunking" of the earlier rape study.

(albeit a mistake likely to be made by exactly the opposite people)

There are certainly things that have been "well-refuted" and "debunked". Andrew Wakefield's study purporting to prove that vaccines cause autism is a pretty good example. But you will notice that it had multiple failed replications, journals published reports showing he falsified data, the study's co-authors retracted their support, the journal it was published in retracted it and issued an apology, the General Medical Council convicted Wakefield of sixteen counts of misconduct, and Wakefield was stripped of his medical license and barred from practicing medicine ever again in the UK. The *British Medical Journal*, one of the best-respected medical journals in the world, published an editorial concluding:

> Clear evidence of falsification of data should now close the door on this damaging vaccine scare … Who perpetrated this fraud? There is no doubt that it was Wakefield. Is it possible that he was wrong, but not dishonest: that he was so incompetent that he was unable to fairly describe the project, or to report even one of the 12 children's cases accurately? No.

Meanwhile, *The Bell Curve* was lambasted in the popular press and by many academics. But it also got fifty of the top

researchers in its field to sign a consensus statement saying it was pretty much right about everything and the people attacking it were biased and confused. Three years later, they re-issued their statement saying nothing had changed and more recent findings had only confirmed their opinion. The American Psychological Association launched a task force to settle the issue which stopped short of complete agreement but which given the circumstances was pretty darned supportive. There are certainly a lot of smart people with very strong negative opinions, but each one is still usually met by an equally ardent and credentialed proponent.

One of these two things has been "well-refuted". The other has been "argued against".

III.

I saw this same dynamic at work the other day, looking through the minimum wage literature.

The primordial titanomachy of the minimum wage literature goes like this. In 1994, two guys named Card and Krueger published a study showing the minimum wage had if anything positive effects on New Jersey restaurants, convincing many people that minimum wages were good. In 1996, two guys named Neumark and Wascher reanalyzed the New Jersey data using a different source and found that it showed the minimum wage had very bad effects on New Jersey restaurants. In 2000, Card and Krueger responded, saying that their analysis was better than Neumark and Wascher's re-analysis, and also they had done a re-analysis of their own which confirmed their original position.

Let's see how conservative sites present this picture:

"The support for this assertion is the oft-cited 1994 study by Card and Krueger showing a positive correlation between an

increased minimum wage and employment in New Jersey. Many others have thoroughly debunked this study." (source)

"I was under the impression that the original study done by Card and Krueger had been thoroughly debunked by Michigan State University economist David Neumark and William Wascher" (source)

"The study … by Card and Krueger has been debunked by several different people several different times. When other researchers re-evaluated the study, they found that data collected using those records 'lead to the opposite conclusion from that reached by' Card and Krueger." (source)

"It was only a short time before the fantastic Card-Krueger findings were challenged and debunked by several subsequent studies…in 1995, economists David Neumark and David Wascher used actual payroll records (instead of survey data used by Card and Krueger) and published their results in an NBER paper with an amazing finding: Demand curves for unskilled labor really do slope downward, confirming 200 years of economic theory and mountains of empirical evidence (source)

And now let's look at how lefty sites present this picture:

"…a long-debunked paper [by Neumark and Wascher]" (source)

"Note that your Mises heroes, Neumark and Wascher are roundly debunked." (source)

"Neumark's living wage and minimum wage research have been found to be seriously flawed…based on faulty methods which when corrected refute his conclusion." – (source)

"…Neumark and Wascher, a study which Elizabeth Warren debunked in a Senate hearing" (source)

So if you're conservative, Neumark and Wascher debunked Card and Krueger. But if you're liberal, Card and Krueger debunked Neumark and Wascher.

Both sides are no doubt very pleased with themselves. They're not men of one study. They look at *all* of the research – except of course the studies that have been "debunked" or "well-refuted". Why would you waste your time with *those?*

IV.

Once again, I'm not preaching radical skepticism.

First of all, some studies are *super-debunked*. Wakefield is a good example.

Second of all, some studies that don't quite meet Wakefield-level of awfulness are indeed really bad and need refuting. I don't think this is beyond the intellectual capacities of most people. I think in many cases it's easy to understand why a study is wrong, you should try to do that, and once you do it you can safely discount the results of the study.

I'm not against pointing out when you disagree with studies or think they're flawed. I'd be a giant hypocrite if I was.

But "debunked" and "refuted" aren't saying you disagree with a study. They're making arguments from authority. They're saying "the authority of the scientific community has come together and said this is a piece of crap that doesn't count".

And that's fine if that's actually happened. But you had better make sure that you're calling upon an ex cathedra statement by the community itself, and not a single guy with an axe to grind. Or one side of a complicated an interminable debate where both sides have about equal credentials and sway.

If you can't do that, you say "I think that my side of the academic debate is in the right, and here's why," not "your

side has been debunked".

Otherwise you're going to end up like the minimum wage debaters, where both sides claim to have debunked the other. Or like that woman on Twitter, who calls a common position backed by leading researchers "well-refuted". Or like the Federalist article that says a study has been "put to bed" as "bogus" just because another study said something different.

I think this is part of my reply to the claim that empiricism is so great that no one needs rationality.

A naive empiricist who swears off critical thinking because they can just "follow the evidence" has no contingency plan for when the evidence gets confusing. Their only recourse is to deny that the evidence is confusing, to assert that one side or the other has been "debunked". Since they've already made a principled decision not to study confirmation bias, chances are it's going to be whichever side they don't like that's "already been debunked". And by "debunked" they mean "a scientist on my side said it was wrong, so now I am relieved from the burden of thinking about it."

On the original post, I wrote:

> Life is made up of limited, confusing, contradictory, and maliciously doctored facts. Anyone who says otherwise is either sticking to such incredibly easy solved problems that they never encounter anything outside their comfort level, or so closed-minded that they shut out any evidence that challenges their beliefs.

In the absence of any actual debunking more damning than a counterargument, "that's been debunked" is the way "shuts out any evidence that challenges their beliefs" feels from the inside.

V.

Somebody's going to want to know what's up with the original rape studies. The answer is that a small part of the discrepancy is response bias on the CSAS, but most of it is that the two surveys encourage respondents to define "sexual assault" in very different ways. Vox has an excellent article on this which for once I 100% endorse.

In other words, both are valid, both come together to form a more nuanced picture of campus violence, and neither one "debunks" the other. How about that?

How to Not Lose an Argument

Related to: Leave a Line of Retreat

Followup to: Talking Snakes: A Cautionary Tale, The Skeptic's Trilemma

> *"I argue very well. Ask any of my remaining friends. I can win an argument on any topic, against any opponent. People know this, and steer clear of me at parties. Often, as a sign of their great respect, they don't even invite me."*
>
> —Dave Barry

The science of winning arguments is called Rhetoric, and it is one of the Dark Arts. Its study is forbidden to rationalists, and its tomes and treatises are kept under lock and key in a particularly dark corner of the Miskatonic University library. More than this it is not lawful to speak.

But I do want to talk about a very closely related skill: not losing arguments.

Rationalists probably find themselves in more arguments than the average person. And if we're doing it right, the truth is hopefully on our side and the argument is ours to lose. And far too often, we *do* lose arguments, even when we're right. Sometimes it's because of biases or inferential distances or other things that can't be helped. But all too often it's because we're shooting ourselves in the foot.

How does one avoid shooting one's self in the foot? In rationalist language, the technique is called Leaving a Social Line of Retreat. In normal language, it's called being nice.

First, what does it mean to win or lose an argument? There is an unspoken belief in some quarters that the point of an

argument is to gain social status by utterly demolishing your opponent's position, thus proving yourself the better thinker. That can be fun sometimes, and if it's really all you want, go for it.

But the most important reason to argue with someone is to change his mind. If you want a world without fundamentalist religion, you're never going to get there just by making cutting and incisive critiques of fundamentalism that all your friends agree sound really smart. You've got to deconvert some actual fundamentalists. In the absence of changing someone's mind, you can at least get them to see your point of view. Getting fundamentalists to understand the real reasons people find atheism attractive is a nice consolation prize.

I make the anecdotal observation that a lot of smart people are very good at winning arguments in the first sense, and very bad at winning arguments in the second sense. Does that correspond to your experience?

Back in 2008, Eliezer described how to Leave a Line of Retreat. If you believe morality is impossible without God, you have a strong disincentive to become an atheist. Even after you've realized which way the evidence points, you'll activate every possible defense mechanism for your religious beliefs. If all the defense mechanisms fail, you'll take God on utter faith or just believe in belief, rather than surrender to the unbearable position of an immoral universe.

The correct procedure for dealing with such a person, Eliezer suggests, isn't to show them yet another reason why God doesn't exist. They'll just reject it along with all the others. The correct procedure is to convince them, on a gut level, that morality is possible even in a godless universe. When disbelief

in God is no longer so terrifying, people won't fight it quite so hard and may even deconvert themselves.

But there's another line of retreat to worry about, one I experienced firsthand in a very strange way. I had a dream once where God came down to Earth; I can't remember exactly why. In the borderlands between waking and sleep, I remember thinking: *I feel like a total moron*. Here I am, someone who goes to atheist groups and posts on atheist blogs and has told all his friends they should be atheists and so on, and now it turns out God exists. All of my religious friends whom I won all those arguments against are going to be secretly looking at me, trying as hard as they can to be nice and understanding, but secretly laughing about how I got my comeuppance. I can never show my face in public again. Wouldn't you feel the same?

And then I woke up, and shook it off. I am an aspiring rationalist: if God existed, I would desire to believe that God existed. But I realized at that point the importance of the social line of retreat. The psychological resistance I felt to admitting God's existence, even after having seen Him descend to Earth, was immense. And, I realized, it was exactly the amount of resistance that every vocally religious person must experience towards God's *non*-existence.

There's not much we can do about this sort of high-grade long-term resistance. Either a person has enough of the rationalist virtues to overcome it, or he doesn't. But there is a less ingrained, more immediate form of social resistance generated with every heated discussion.

Let's say you approach a theist (let's call him Theo) and say "How can you, a grown man, still believe in something stupid like talking snakes and magic sky kings? Don't you know you

people are responsible for the Crusades and the Thirty Years' War and the Spanish Inquisition? You should be ashamed of yourself!"

This suggests the following dichotomy in Theo's mind: **EITHER** God exists, **OR** I am an idiot who believes in stupid childish things and am in some way partly responsible for millions of deaths and I should have lower status and this arrogant person who's just accosted me and whom I already hate should have higher status at my expense.

Unless Theo has attained a level of rationality far beyond any of us, guess which side of that dichotomy he's going to choose? In fact, guess which side of that dichotomy he's now going to support with renewed vigor, even if he was only a lukewarm theist before? His social line of retreat has been completely closed off, and it's *your* fault.

Here the two definitions of "winning an argument" I suggested before come into conflict. If your goal is to absolutely demolish the other person's position, to make him feel awful and worthless - then you are also very unlikely to change his mind or win his understanding. And because our culture of debates and mock trials and real trials and flaming people on Usenet encourages the first type of "winning an argument", there's precious little genuine mind-changing going on.

Really adjusting to the second type of argument, where you try to convince people, takes a lot more than just not insulting people outright[1]. You've got to completely rethink your entire strategy. For example, anyone used to the Standard Debates may already have a cached pattern of how they work. Activate the whole Standard Debate concept, and you activate a whole bunch of related thoughts like Atheists As The Enemy, Defending The Faith, and even in some cases (I've seen it

happen) persecution of Christians by atheists in Communist Russia. To such a person, ceding an inch of ground in a Standard Debate may well be equivalent to saying all the Christians martyred by the Communists died in vain, or something similarly dreadful.

So try to show you're not just starting Standard Debate #4457. I remember once, during the middle of a discussion with a Christian, when I admitted I really didn't like Christopher Hitchens. Richard Dawkins, brilliant. Daniel Dennett, brilliant. But Christopher Hitchens always struck me as too black-and-white and just plain irritating. This one little revelation completely changed the entire tone of the conversation. I was no longer Angry Nonbeliever #116. I was no longer the living incarnation of All Things Atheist. I was just a person who happened to have a whole bunch of atheist ideas, along with a couple of ideas that weren't typical of atheists. I got the same sort of response by admitting I loved religious music. All of a sudden my friend was falling over himself to mention some scientific theory he found especially elegant in order to reciprocate[2]. I didn't end up deconverting him on the spot, but think he left with a much better appreciation of my position.

All of these techniques fall dangerously close to the Dark Arts, so let me be clear: I'm not suggesting you misrepresent yourself just to win arguments. I don't think misrepresenting yourself would even work; evolutionary psychology tells us humans are notoriously bad liars. Don't fake an appreciation for the other person's point of view, actually *develop* an appreciation for the other person's point of view. Realize that your points probably seem as absurd to others as their points seem to you. Understand that many false beliefs don't come from simple lying or stupidity, but from complex mixtures of truth and falsehood filtered by complex cognitive biases.

Don’t stop believing that you are right and they are wrong, unless the evidence points that way. But leave it at them being wrong, not them being wrong and stupid and evil.

I think most people intuitively understand this. But considering how many smart people I see shooting their own foot off when they’re trying to convince someone[3], some of them clearly need a reminder.

Footnotes

1: An excellent collection of the deeper and most subtle forms of this practice of this sort can be found in Dale Carnegie’s How to Win Friends and Influence People, one of the only self-help books I’ve read that was truly useful and not a regurgitation of cliches and applause lights. Carnegie’s thesis is basically that being nice is the most powerful of the Dark Arts, and that a master of the Art of Niceness can use it to take over the world. It works better than you’d think.

2: The following technique is definitely one of the Dark Arts, but I mention it because it reveals a lot about the way we think: when engaged in a really heated, angry debate, one where the insults are flying, suddenly stop and admit the other person is one hundred percent right and you’re sorry for not realizing it earlier. Do it properly, and the other person will be flabbergasted, and feel deeply guilty at all the names and bad feelings they piled on top of you. Not only will you ruin their whole day, but for the rest of time, this person will secretly feel indebted to you, and you will be able to play with their mind in all sorts of little ways.

3: Libertarians, you have a particular problem with this. If I wanted to know why I’m a Stalin-worshipper who has betrayed the Founding Fathers for personal gain and is

controlled by his base emotions and wants to dominate others by force to hide his own worthlessness et cetera, I'd ask Ann Coulter. You're better than that. Come on. And then you wonder why people never vote for you.

The Least Convenient Possible World

Related to: Is That Your True Rejection?

> *"If you're interested in being on the right side of disputes, you will refute your opponents' arguments. But if you're interested in producing truth, you will fix your opponents' arguments for them. To win, you must fight not only the creature you encounter; you must fight the most horrible thing that can be constructed from its corpse."*
>
> — Black Belt Bayesian, via Rationality Quotes 13

Yesterday John Maxwell's post wondered how much the average person would do to save ten people from a ruthless tyrant. I remember asking some of my friends a vaguely related question as part of an investigation of the Trolley Problems:

> You are a doctor in a small rural hospital. You have ten patients, each of whom is dying for the lack of a separate organ; that is, one person needs a heart transplant, another needs a lung transplant, another needs a kidney transplant, and so on. A traveller walks into the hospital, mentioning how he has no family and no one knows that he's there. All of his organs seem healthy. You realize that by killing this traveller and distributing his organs among your patients, you could save ten lives. Would this be moral or not?

I don't want to discuss the answer to this problem today. I want to discuss the answer one of my friends gave, because I think it illuminates a very interesting kind of defense

mechanism that rationalists need to be watching for. My friend said:

> It wouldn't be moral. After all, people often reject organs from random donors. The traveller would probably be a genetic mismatch for your patients, and the transplantees would have to spend the rest of their lives on immunosuppressants, only to die within a few years when the drugs failed.

On the one hand, I have to give my friend credit: his answer is biologically accurate, and beyond a doubt the technically correct answer to the question I asked. On the other hand, I don't have to give him very *much* credit: he completely missed the point and lost a valuable effort to examine the nature of morality.

So I asked him, "In the least convenient possible world, the one where everyone was genetically compatible with everyone else and this objection was invalid, what would you do?"

He mumbled something about counterfactuals and refused to answer. But I learned something very important from him, and that is to always ask this question of *myself*. Sometimes the least convenient possible world is the only place where I can figure out my true motivations, or which step to take next. I offer three examples:

1: Pascal's Wager. Upon being presented with Pascal's Wager, one of the first things most atheists think of is this:

Perhaps God values intellectual integrity so highly that He is prepared to reward honest atheists, but will punish anyone who practices a religion he does not truly believe simply for personal gain. Or perhaps, as the Discordians claim, "Hell is

reserved for people who believe in it, and the hottest levels of Hell are reserved for people who believe in it on the principle that they'll go there if they don't."

This is a good argument against Pascal's Wager, but it isn't the least convenient possible world. The least convenient possible world is the one where Omega, the completely trustworthy superintelligence who is always right, informs you that God definitely doesn't value intellectual integrity that much. In fact (Omega tells you) either God does not exist or the Catholics are right about absolutely everything.

Would you become a Catholic in this world? Or are you willing to admit that maybe your rejection of Pascal's Wager has less to do with a hypothesized pro-atheism God, and more to do with a belief that it's wrong to abandon your intellectual integrity on the off chance that a crazy deity is playing a perverted game of blind poker with your eternal soul?

2: The God-Shaped Hole. Christians claim there is one in every atheist, keeping him from spiritual fulfillment.

Some commenters on Raising the Sanity Waterline don't deny the existence of such a hole, if it is intepreted as a desire for purpose or connection to something greater than one's self. But, some commenters say, science and rationality can fill this hole even better than God can.

What luck! Evolution has by a wild coincidence created us with a big rationality-shaped hole in our brains! Good thing we happen to be rationalists, so we can fill this hole in the best possible way! I don't know - despite my sarcasm this may even be true. But in the least convenient possible world, Omega comes along and tells you that sorry, the hole is exactly God-shaped, and anyone without a religion will lead a less-than-optimally-happy life. Do you head down to the

nearest church for a baptism? Or do you admit that even if believing something makes you happier, you still don't want to believe it unless it's true?

3: **Extreme Altruism.** John Maxwell mentions the utilitarian argument for donating almost everything to charity.

Some commenters object that many forms of charity, especially the classic "give to starving African orphans," are counterproductive, either because they enable dictators or thwart the free market. This is quite true.

But in the least convenient possible world, here comes Omega again and tells you that Charity X has been proven to do exactly what it claims: help the poor without any counterproductive effects. So is your real objection the corruption, or do you just not believe that you're morally obligated to give everything you own to starving Africans?

You may argue that this citing of convenient facts is at worst a venial sin. If you still get to the correct answer, and you do it by a correct method, what does it matter if this method isn't really the one that's convinced you personally?

One easy answer is that it saves you from embarrassment later. If some scientist does a study and finds that people really do have a god-shaped hole that can't be filled by anything else, no one can come up to you and say "Hey, didn't you say the reason you didn't convert to religion was because rationality filled the god-shaped hole better than God did? Well, I have some bad news for you…"

Another easy answer is that your real answer teaches you something about yourself. My friend may have successfully avoiding making a distasteful moral judgment, but he didn't

learn anything about morality. My refusal to take the easy way out on the transplant question helped me develop the form of precedent-utilitarianism I use today.

But more than either of these, it matters because it seriously influences where you go next.

Say "I accept the argument that I need to donate almost all my money to poor African countries, but my only objection is that corrupt warlords might get it instead", and the *obvious* next step is to see if there's a poor African country without corrupt warlords (see: Ghana, Botswana, etc.) and donate almost all your money to them. Another acceptable answer would be to donate to another warlord-free charitable cause like the Singularity Institute.

If you *just* say "Nope, corrupt dictators might get it," you may go off and spend the money on a new TV. Which is fine, *if* a new TV is what you really want. But if you're the sort of person who *would have* been convinced by John Maxwell's argument, but you dismissed it by saying "Nope, corrupt dictators," then you've lost an opportunity to change your mind.

So I recommend: limit yourself to responses of the form "I completely reject the entire basis of your argument" or "I accept the basis of your argument, but it doesn't apply to the real world because of contingent fact X." If you just say "Yeah, well, contigent fact X!" and walk away, you've left yourself too much wiggle room.

In other words: always have a plan for what you would do in the least convenient possible world.

Bayes for Schizophrenics: Reasoning in Delusional Disorders

Related to: The Apologist and the Revolutionary, Dreams with Damaged Priors

Several years ago, I posted about V.S. Ramachandran's 1996 theory explaining anosognosia through an "apologist" and a "revolutionary".

Anosognosia, a condition in which extremely sick patients mysteriously deny their sickness, occurs during right-sided brain injury but not left-sided brain injury. It can be extraordinarily strange: for example, in one case, a woman whose left arm was paralyzed insisted she could move her left arm just fine, and when her doctor pointed out her immobile arm, she claimed that was her daughter's arm even though it was obviously attached to her own shoulder. Anosognosia can be temporarily alleviated by squirting cold water into the patient's left ear canal, after which the patient suddenly realizes her condition but later loses awareness again and reverts back to the bizarre excuses and confabulations.

Ramachandran suggested that the left brain is an "apologist", trying to justify existing theories, and the right brain is a "revolutionary" which changes existing theories when conditions warrant. If the right brain is damaged, patients are unable to change their beliefs; so when a patient's arm works fine until a right-brain stroke, the patient cannot discard the hypothesis that their arm is functional, and can only use the left brain to try to fit the facts to their belief.

In the almost twenty years since Ramachandran's theory was published, new research has kept some of the general outline

while changing many of the specifics in the hopes of explaining a wider range of delusions in neurological and psychiatric patients. The newer model acknowledges the left-brain/right-brain divide, but adds some new twists based on the Mind Projection Fallacy and the brain as a Bayesian reasoner.

INTRODUCTION TO DELUSIONS

Strange as anosognosia is, it's only one of several types of delusions, which are broadly categorized into polythematic and monothematic. Patients with polythematic delusions have multiple unconnected odd ideas: for example, the famous schizophrenic game theorist John Nash believed that he was defending the Earth from alien attack, that he was the Emperor of Antarctica, *and* that he was the left foot of God. A patient with a monothematic delusion, on the other hand, usually only has one odd idea. Monothematic delusions vary less than polythematic ones: there are a few that are relatively common across multiple patients. For example:

In the Capgras delusion, the patient, usually a victim of brain injury but sometimes a schizophrenic, believes that one or more people close to her has been replaced by an identical imposter. For example, one male patient expressed the worry that his wife was actually someone else, who had somehow contrived to exactly copy his wife's appearance and mannerisms. This delusion sounds harmlessly hilarious, but it can get very ugly: in at least one case, a patient got so upset with the deceit that he murdered the hypothesized imposter - actually his wife.

The Fregoli delusion is the opposite: here the patient thinks that random strangers she meets are actually her friends and

family members in disguise. Sometimes everyone may be the same person, who must be as masterful at quickly changing costumes as the famous Italian actor Fregoli (inspiring the condition's name).

In the Cotard delusion, the patient believes she is dead. Cotard patients will neglect personal hygiene, social relationships, and planning for the future - as the dead have no need to worry about such things. Occasionally they will be able to describe in detail the "decomposition" they believe they are undergoing.

Patients with all these types of delusions[1] - as well as anosognosiacs - share a common feature: they usually have damage to the right frontal lobe of the brain (including in schizophrenia, where the brain damage is of unknown origin and usually generalized, but where it is still possible to analyze which areas are the most abnormal). It would be nice if a theory of anosognosia also offered us a place to start explaining these other conditions, but this Ramachandran's idea fails to do. He posits a problem with belief shift: going from the originally correct but now obsolete "my arm is healthy" to the updated "my arm is paralyzed". But these other delusions cannot be explained by simple failure to update: delusions like "the person who appears to be my wife is an identical imposter" *never* made sense. We will have to look harder.

ABNORMAL PERCEPTION: THE FIRST FACTOR

Coltheart, Langdon, and McKay posit what they call the "two-factor theory" of delusion. In the two-factor theory, one problem causes an abnormal perception, and a second problem causes the brain to come up with a bizarre instead of a reasonable explanation.

Abnormal perception has been best studied in the Capgras delusion. A series of experiments, including some by Ramachandran himself, demonstrate that Capgras patients lack a skin conductance response (usually used as a proxy of emotional reaction) to familiar faces. This meshes nicely with the brain damage pattern in Capgras, which seems to involve the connection between the face recognition areas in the temporal lobe and the emotional areas in the limibic system. So although the patient can recognize faces, and can feel emotions, the patient cannot feel emotions related to recognizing faces.

The older "one-factor" theories of delusion stopped here. The patient, they said, knows that his wife looks like his wife, but he doesn't feel any emotional reaction to her. If it was really his wife, he would feel something - love, irritation, whatever - but he feels only the same blankness that would accompany seeing a stranger. Therefore (the one-factor theory says) his brain gropes for an explanation and decides that she really is a stranger. Why does this stranger look like his wife? Well, she must be wearing a very good disguise.

One-factor theories also do a pretty good job of explaining many of the remaining monothematic delusions. A 1998 experiment shows that Cotard delusion sufferers have a globally decreased autonomic response: that is, nothing really makes them feel much of anything - a state consistent with being dead. And anosognosiacs have lost not only the nerve connections that would allow them to move their limbs, but the nerve connections that would send distress signals and even the connections that would send back "error messages" if the limb failed to move correctly - so the brain gets data that everything is fine.

The basic principle behind the first factor is "Assume that reality is such that my mental states are justified", a sort of Super Mind Projection Fallacy.

Although I have yet to find an official paper that says so, I think this same principle also explains many of the more typical schizophrenic delusions, of which two of the most common are delusions of grandeur and delusions of persecution. Delusions of grandeur are the belief that one is extremely important. In pop culture, they are typified by the psychiatric patient who believes he is Jesus or Napoleon - I've never met any Napoleons, but I know several Jesuses and recently worked with a man who thought he was Jesus and John Lennon at the same time. Here the first factor is probably an elevated mood (working through a miscalibrated sociometer). "Wow, I feel like I'm really awesome. In what case would I be justified in thinking so highly of myself? Only if I were Jesus and John Lennon at the same time!" A similar mechanism explains delusions of persecution, the classic "the CIA is after me" form of disease. We apply the Super Mind Projection Fallacy to a garden-variety anxiety disorder: "In what case would I be justified in feeling this anxious? Only if people were constantly watching me and plotting to kill me. Who could do that? The CIA."

But despite the explanatory power of the Super Mind Projection Fallacy, the one-factor model isn't enough.

ABNORMAL BELIEF EVALUATION: THE SECOND FACTOR

The one-factor model requires people to be really stupid. Many Capgras patients were normal intelligent people before their injuries. Surely they wouldn't leap straight from "I don't feel affection when I see my wife's face" to "And therefore

this is a stranger who has managed to look exactly like my wife, sounds exactly like my wife, owns my wife's clothes and wedding ring and so on, and knows enough of my wife's secrets to answer any question I put to her exactly like my wife would." The lack of affection vaguely supports the stranger hypothesis, but the prior for the stranger hypothesis is so low that it should never even enter consideration (remember this phrasing: it will become important later.) Likewise, we've all felt really awesome at one point or another, but it's never occurred to most of us that maybe we are simultaneously Jesus and John Lennon.

Further, most psychiatric patients with the deficits involved don't develop delusions. People with damage to the ventromedial area suffer the same disconnection between face recognition and emotional processing as Capgras patients, but they don't draw any unreasonable conclusions from it. Most people who get paralyzed don't come down with anosognosia, and most people with mania or anxiety don't think they're Jesus or persecuted by the CIA. What's the difference between these people and the delusional patients?

The difference is the right dorsolateral prefrontal cortex, an area of the brain strongly associated with delusions. If whatever brain damage broke your emotional reactions to faces or paralyzed you or whatever spared the RDPC, you are unlikely to develop delusions. If your brain damage also damaged this area, you are correspondingly more likely to come up with a weird explanation.

In his first papers on the subject, Coltheart vaguely refers to the RDPC as a "belief evaluation" center. Later, he gets more specific and talks about its role in Bayesian updating. In his chronology, a person damages the connection between face recognition and emotion, and "rationally" concludes the

Capgras hypothesis. In his model, even if there's only a 1% prior of your spouse being an imposter, if there's a 1000 times greater likelihood of you not feeling anything toward an imposter than to your real spouse, you can "rationally" come to believe in the delusion. In normal people, this rational belief then gets worn away by updating based on evidence: the imposter seems to know your spouse's personal details, her secrets, her email passwords. In most patients, this is sufficient to have them update back to the idea that it is really their spouse. In Capgras patients, the damage to the RDPC prevents updating on "exogenous evidence" (for some reason, the endogenous evidence of the lack of emotion itself still gets through) and so they maintain their delusion.

This theory has some trouble explaining why patients are still able to update about other situations, but Coltheart speculates that maybe the belief evaluation system is weakened but not totally broken, and can deal with anything except the ceaseless stream of contradictory endogenous information.

EXPLANATORY ADEQUACY BIAS

McKay makes an excellent critique of several questionable assumptions of this theory.

First, is the Capgras hypothesis *ever* plausible? Coltheart et al pretend that the prior is 1/100, but this implies that there is a base rate of your spouse being an imposter one out of every hundred times you see her (or perhaps one out of every hundred people has a fake spouse) either of which is preposterous. No reasonable person could entertain the Capgras hypothesis even for a second, let alone for long enough that it becomes their working hypothesis and develops immunity to further updating from the broken RDPC.

Second, there's no evidence that the ventromedial patients - the ones who lose face-related emotions but don't develop the Capgras delusion - once had the Capgras delusion but then successfully updated their way out of it. They just never develop the delusion to begin with.

McKay keeps the Bayesian model, but for him the second factor is not a deficit in updating in general, but a deficit in the use of priors. He lists two important criteria for reasonable belief: "explanatory adequacy" (what standard Bayesians call the likelihood ratio; the new data must be more likely if the new belief is true than if it is false) and "doxastic conservativism" (what standard Bayesians call the prior; the new belief must be reasonably likely to begin with given everything else the patient knows about the world).

Delusional patients with damage to their RDPC lose their ability to work with priors and so abandon all doxastic conservativism, essentially falling into a what we might term the Super Base Rate Fallacy. For them the only important criterion for a belief is explanatory adequacy. So when they notice their spouse's face no longer elicits any emotion, they decide that their spouse is not really their spouse at all. This does a great job of explaining the observed data - maybe the best job it's possible for an explanation to do. Its only minor problem is that it has a stupendously low prior, and this doesn't matter because they are no longer able to take priors into account.

This also explains why the delusional belief is impervious to new evidence. Suppose the patient's spouse tells personal details of their honeymoon that no one else could possibly know. There are several possible explanations: the patient's spouse really is the patient's spouse, or (says the left-brain Apologist) the patient's spouse is an alien who was able to

telepathically extract the relevant details from the patient's mind. The telepathic alien imposter hypothesis has great explanatory adequacy: it explains why the person looks like the spouse (the alien is a very good imposter), why the spouse produces no emotional response (it's not the spouse at all) and why the spouse knows the details of the honeymoon (the alien is telepathic). The "it's really your spouse" explanation only explains the first and the third observations. Of course, we as sane people know that the telepathic alien hypothesis has a very low base rate plausibility because of its high complexity and violation of Occam's Razor, but these are exactly the factors that the RDPC-damaged[2] patient can't take into account. Therefore, the seemingly convincing new evidence of the spouse's apparent memories only suffices to help the delusional patient infer that the imposter is telepathic.

The Super Base Rate Fallacy can explain the other delusional states as well. I recently met a patient who was, indeed, convinced the CIA were after her; of note she also had extreme anxiety to the point where her arms were constantly shaking and she was hiding under the covers of her bed. CIA pursuit is probably the best possible reason to be anxious; the only reason we don't use it more often is how few people are really pursued by the CIA (well, as far as we know). My mentor warned me not to try to argue with the patient or convince her that the CIA wasn't really after her, as (she said from long experience) it would just make her think I was in on the conspiracy. This makes sense. "The CIA is after you and your doctor is in on it" explains both anxiety and the doctor's denial of the CIA very well; "The CIA is not after you" explains only the doctor's denial of the CIA. For anyone with a pathological inability to handle Occam's Razor, the best solution to a challenge to your hypothesis is always to make

your hypothesis more elaborate.

OPEN QUESTIONS

Although I think McKay's model is a serious improvement over its predecessors, there are a few loose ends that continue to bother me.

"You have brain damage" is also a theory with perfect explanatory adequacy. If one were to explain the Capgras delusion to Capgras patients, it would provide just as good an explanation for their odd reactions as the imposter hypothesis. Although the patient might not be able to appreciate its decreased complexity, they should at least remain indifferent between the two hypotheses. I've never read of any formal study of this, but given that someone must have tried explaining the Capgras delusion to Capgras patients I'm going to assume it doesn't work. Why not?

Likewise, how come delusions are so specific? It's impossible to convince someone who thinks he is Napoleon that he's really just a random non-famous mental patient, but it's also impossible to convince him he's Alexander the Great (at least I think so; I don't know if it's ever been tried). But him being Alexander the Great is also consistent with his observed data and his deranged inference abilities. Why decide it's the CIA who's after you, and not the KGB or Bavarian Illuminati?

Why is the failure so often limited to failed inference from mental states? That is, if a Capgras patient sees it is raining outside, the same process of base rate avoidance that made her fall for the Capgras delusion ought to make her think she's been transported to ther rainforest or something. This happens in polythematic delusion patients, where anything at all can generate a new delusion, but not those with monothematic

delusions like Capgras. There must be some fundamental difference between how one draws inferences from mental states versus everything else.

This work also raises the question of whether one can one consciously use System II Bayesian reasoning to argue oneself out of a delusion. It seems improbable, but I recently heard about an n=1 personal experiment of a rationalist with schizophrenia who used successfully used Bayes to convince themselves that a delusion (or possibly hallucination; the story was unclear) was false. I don't have their permission to post their story here, but I hope they'll appear in the comments.

FOOTNOTES

1: I left out discussion of the Alien Hand Syndrome, even though it was in my sources, because I believe it's more complicated than a simple delusion. There's some evidence that the alien hand actually does move independently; for example it will sometimes attempt to thwart tasks that the patient performs voluntarily with their good hand. Some sort of "split brain" issues seem like a better explanation than simple Mind Projection.

2: The right dorsolateral prefrontal cortex also shows up in dream research, where it tends to be one of the parts of the brain shut down during dreaming. This provides a reasonable explanation of why we don't notice our dreams' implausibility while we're dreaming them - and Eliezer specifically mentions he can't use priors correctly in his dreams. It also highlights some interesting parallels between dreams and the monothematic delusions. For example, the typical "And then I saw my mother, but she was also somehow my fourth grade teacher at the same time" effect seems sort of like Capgras and Fregoli. Even more interestingly, the RDPC gets switched on

during lucid dreaming, providing an explanation of why lucid dreamers are able to reason normally in dreams. Because lucid dreaming also involves a sudden “switching on” of “awareness”, this makes the RDPC a good target area for consciousness research.

Generalizing from One Example

Related to: The Psychological Unity of Humankind, Instrumental vs. Epistemic: A Bardic Perspective

"*Everyone generalizes from one example. At least, I do.*"

— Vlad Taltos (*Issola,* Steven Brust)

My old professor, David Berman, liked to talk about what he called the "typical mind fallacy", which he illustrated through the following example:

There was a debate, in the late 1800s, about whether "imagination" was simply a turn of phrase or a real phenomenon. That is, can people actually create images in their minds which they see vividly, or do they simply say "I saw it in my mind" as a metaphor for considering what it looked like?

Upon hearing this, my response was "How the stars was this actually a real debate? Of course we have mental imagery. Anyone who doesn't think we have mental imagery is either such a fanatical Behaviorist that she doubts the evidence of her own senses, or simply insane." Unfortunately, the professor was able to parade a long list of famous people who denied mental imagery, including some leading scientists of the era. And this was all before Behaviorism even existed.

The debate was resolved by Francis Galton, a fascinating man who among other achievements invented eugenics, the "wisdom of crowds", and standard deviation. Galton gave people some very detailed surveys, and found that some people did have mental imagery and others didn't. The ones who did had simply assumed everyone did, and the ones who didn't had simply assumed everyone didn't, to the point of

coming up with absurd justifications for why they were lying or misunderstanding the question. There was a wide spectrum of imaging ability, from about five percent of people with perfect eidetic imagery[1] to three percent of people completely unable to form mental images[2].

Dr. Berman dubbed this the Typical Mind Fallacy: the human tendency to believe that one's own mental structure can be generalized to apply to everyone else's.

He kind of took this idea and ran with it. He interpreted certain passages in George Berkeley's biography to mean that Berkeley was an eidetic imager, and that this was why the idea of the universe as sense-perception held such interest to him. He also suggested that experience of consciousness and qualia were as variable as imaging, and that philosophers who deny their existence (Ryle? Dennett? Behaviorists?) were simply people whose mind lacked the ability to easily experience qualia. In general, he believed philosophy of mind was littered with examples of philosophers taking their own mental experiences and building theories on them, and other philosophers with different mental experiences critiquing them and wondering why they disagreed.

The formal typical mind fallacy is about serious matters of mental structure. But I've also run into something similar with something more like the psyche than the mind: a tendency to generalize from our personalities and behaviors.

For example, I'm about as introverted a person as you're ever likely to meet - anyone more introverted than I am doesn't communicate with anyone. All through elementary and middle school, I suspected that the other children were out to get me. They kept on grabbing me when I was busy with something and trying to drag me off to do some rough activity with them

and their friends. When I protested, they counter-protested and told me I really needed to stop whatever I was doing and come join them. I figured they were bullies who were trying to annoy me, and found ways to hide from them and scare them off.

Eventually I realized that it was a double misunderstanding. They figured I must be like them, and the only thing keeping me from playing their fun games was that I was too shy. I figured they must be like me, and that the only reason they would interrupt a person who was obviously busy reading was that they wanted to annoy him.

Likewise: I can't deal with noise. If someone's being loud, I can't sleep, I can't study, I can't concentrate, I can't do anything except bang my head against the wall and hope they stop. I once had a noisy housemate. Whenever I asked her to keep it down, she told me I was being oversensitive and should just mellow out. I can't claim total victory here, because she was very neat and kept yelling at me for leaving things out of place, and I told her she needed to just mellow out and you couldn't even tell that there was dust on that dresser anyway. It didn't occur to me then that neatness to her might be as necessary and uncompromisable as quiet was to me, and that this was an actual feature of how our minds processed information rather than just some weird quirk on her part.

"Just some weird quirk on her part" and "just being oversensitive" are representative of the problem with the typical psyche fallacy, which is that it's invisible. We tend to neglect the role of differently-built minds in disagreements, and attribute the problems to the other side being deliberately perverse or confused. I happen to know that loud noise seriously pains and debilitates me, but when I say this to other

people they think I'm just expressing some weird personal preference for quiet. Think about all those poor non-imagers who thought everyone else was just taking a metaphor about seeing mental images *way* too far and refusing to give it up.

And the reason I'm posting this here is because it's rationality that helps us deal with these problems.

There's some evidence that the usual method of interacting with people involves something sorta like emulating them within our own brain. We think about how we would react, adjust for the other person's differences, and then assume the other person would react that way. This method of interaction is very tempting, and it always feels like it ought to work.

But when statistics tell you that the method that would work on you doesn't work on anyone else, then continuing to follow that gut feeling is a Typical Psyche Fallacy. You've got to be a good rationalist, reject your gut feeling, and follow the data.

I only really discovered this in my last job as a school teacher. There's a lot of data on teaching methods that students enjoy and learn from. I had some of these methods…inflicted…on me during my school days, and I had no intention of abusing my own students in the same way. And when I tried the sorts of really creative stuff I would have loved as a student…it fell completely flat. What ended up working? Something pretty close to the teaching methods I'd hated as a kid. Oh. Well. Now I know why people use them so much. And here I'd gone through life thinking my teachers were just inexplicably bad at what they did, never figuring out that I was just the odd outlier who couldn't be reached by this sort of stuff.

The other reason I'm posting this here is because I think it relates to some of the discussions of seduction that are going on in MBlume's Bardic thread. There are a lot of not-

particularly-complimentary things about women that many men tend to believe. Some guys say that women will never have romantic relationships with their actually-decent-people male friends because they prefer alpha-male jerks who treat them poorly. Other guys say women want to be lied to and tricked. I could go on, but I think most of them are covered in that thread anyway.

The response I hear from most of the women I know is that this is complete balderdash and women aren't like that at all. So what's going on?

Well, I'm afraid I kind of trust the seduction people. They've put a lot of work into their "art" and at least according to their self-report are pretty successful. And unhappy romantically frustrated nice guys everywhere can't be completely wrong.

My theory is that the women in this case are committing a Typical Psyche Fallacy. The women I ask about this are not even remotely close to being a representative sample of all women. They're the kind of women whom a shy and somewhat geeky guy knows and talks about psychology with. Likewise, the type of women who publish strong opinions about this on the Internet aren't close to a representative sample. They're well-educated women who have strong opinions about gender issues and post about them on blogs.

And lest I sound chauvinistic, the same is certainly true of men. I hear a lot of bad things said about men (especially with reference to what they want romantically) that I wouldn't dream of applying to myself, my close friends, or to any man I know. But they're so common and so well-supported that I have excellent reason to believe they're true.

This post has gradually been getting less rigorous and less connected to the formal Typical Mind Fallacy. First I changed

it to a Typical Psyche Fallacy so I could talk about things that were more psychological and social than mental. And now it's expanding to cover the related fallacy of believing your own social circle is at least a little representative of society at large, which it very rarely is[3].

It was originally titled "The Typical Mind Fallacy", but I'm taking a hint fromt the quote and changing it to "Generalizing From One Example", because that seems to be the link between all of these errors. We only have direct first-person knowledge one one mind, one psyche, and one social circle, and we find it tempting to treat it as typical even in the face of contrary evidence.

This, I think, is especially important for the sort of people who enjoy Less Wrong, who as far as I can tell are with few exceptions the sort of people who are extreme outliers on every psychometric test ever invented.

Footnotes

1. Eidetic imagery, vaguely related to the idea of a "photographic memory", is the ability to visualize something and have it be exactly as clear, vivid and obvious as actually seeing it. My professor's example (which Michael Howard somehow remembers even though I only mentioned it once a few years ago) is that although many people can imagine a picture of a tiger, only an eidetic imager would be able to count the number of stripes.

2. According to Galton, people incapable of forming images were overrepresented in math and science. I've since heard that this idea has been challenged, but I can't access the study.

3. The example that really drove this home to me: what percent of high school students do you think cheat on tests?

What percent have shoplifted? Someone did a survey on this recently and found that the answer was nobhg gjb guveqf unir purngrq naq nobhg bar guveq unir fubcyvsgrq (rot13ed so you have to actually take a guess first). This shocked me and everyone I knew, because we didn't cheat or steal during high school and we didn't know anyone who did. I spent an afternoon trying to find some proof that the study was wrong or unrepresentative and coming up with nothing.

Typical Mind and Politics

Yesterday, in the The Terrible, Horrible, No Good Truth About Morality, Roko mentioned some good evidence that we develop an opinion first based on intuitions, and only later look for rational justifications. For example, people would claim incest was wrong because of worries like genetic defects or later harm, but continue to insist that incest was wrong even after all those worries had been taken away.

Roko's examples take advantage of universal human feelings like the incest taboo. But if people started out with opposite intuitions, then this same mechanism would produce opinions that people hold very strongly and are happy to support with as many reasons and facts as you please, but which are highly resistant to real debate or to contradicting evidence.

Sound familiar?

But to explain politics with this mechanism, we'd need an explanation for why people's intuitions differed to begin with. We've already discussed some such explanations - self-serving biases, influence from family and community, et cetera - but today I want to talk about another possibility.

A few weeks back, I was discussing harms with Bill Swift on Overcoming Bias. In particular, I was arguing that one situation in which there was an open-and-shut case for government restriction of private activity on private property was nuisance noise. I argued that if you were making noise on your property, and I could hear it on my property, that I was being harmed by your actions and that there was clearly just as much a case for government intervention here as if you were firing flaming arrows at me from your property. I fully

expected Bill to agree that this was obviously true but to have some reason why he didn't think it applied to our particular disagreement.

Instead, to my absolute astonishment, Bill said that noise wasn't really a problem. He said he lived on a noisy property and had just stopped whining and gotten on with his life. I didn't really know how to react to this[1], and ended up assuming either that he'd never lived in a really noisy place like I have, or that he was such a blighted ideologue that he was willing to completely contradict common sense in order to preserve his silly argument.

In other words, I was assuming the person I was debating was either astonishingly stupid or willfully evil. And when my thoughts tend in that direction, it usually means I'm missing something.

Luckily in this case I'd already written a long essay explaining my mistake in detail. In Generalizing From One Example, I warned people against assuming everyone's mind is built the same way their own mind is. One particular example I gave was:

> I can't deal with noise. If someone's being loud, I can't sleep, I can't study, I can't concentrate, I can't do anything except bang my head against the wall and hope they stop. I once had a noisy housemate. Whenever I asked her to keep it down, she told me I was being oversensitive and should just mellow out.

So it seems possible to me that I have an oversensitivity to noise and Bill has an undersensitivity to it. When someone around me is being noisy, my intuitions tell me this is

extremely bad and needs to be stopped by any means necessary. And maybe Bill's intuitions tell him that this is a minor non-problem. I won't say that this is actually behind our disagreement on the issue - my guess is that Bill and I would disagree about government regulation of pollution from a factory as well - but I think it contributes and it makes our debate much less productive than it would have been otherwise.

Let me give an example of one place I think a mind difference *is* behind a political opinion. In Money, The Unit of Caring, Eliezer complained that people were too willing to donate time to charity, and too unwilling to donate money to charity. He gave the example of his own experience, where he felt terrible every time he gave away money, but didn't mind a time committment nearly as much. I fired back a response that this was completely foreign to me, because I am happy to give money to charity and often do it before I've even fully thought about what I'm doing, but will groan and make excuses whenever I'm asked to give away time. I also mentioned that this was a general tendency of mine: I have minimal aversion to monetary loss[2], but wasting time makes me angry.

A few months ago, Barack Obama proposed a plan (which he later decided against) to make every high school and college student volunteer a certain amount of time to charity. Although I usually like Obama, I wrote an absolutely scathing essay about how unbearably bad a policy this was. It was a good essay, it convinced a number of people, and I still agree with most of the points in it. But…

…it was completely out of character for me. I'm the sort of person who heckles libertarians with "Stop whining and just pay your damn taxes!" Although I acknowledge that many government policies are inefficient, I tend to just note

“Hmmm, that government policy is suboptimal, it would be an interesting mental puzzle to figure out how to fix it” rather than actually getting angry about it. This Obama proposal was kind of unique in the amount of antipathy it got from me.

So here’s my theory. My brain is organized in such a way that I get minimal negative feelings at the idea of money being taken away from me. We can even localize this anatomically - studies show that the insula is the part responsible for sending a pain signal whenever the loss of money is considered. So let’s say I have a less-than-normally-active insula in this case. And I get a stronger than normal pain signal from wasted time. This explains why I prefer to donate money than time to my favorite charity.

And it could also explain why I’m not a libertarian. One consequence of libertarianism is that you have every right to feel angry when you’re taxed. But I don’t feel angry, so the part of my brain that comes up with rational justifications for my feelings doesn’t need to come up with a rational justification for why taxation is wrong. I do feel angry about being made to do extra work, so my brain adopted libertarian-type arguments in response to the community service proposal. I predict that if I lived in one of those feudal countries with a work levy rather than a tax, I’d be a libertarian, at least until the local knight heard my opinions and cut off my head.

And I don’t mean to pick on libertarians. I know different people have completely different emotional responses to the idea of other people suffering. For example, I can’t watch documentaries on (say) the awful lives on mine workers, because they make me too upset. Other people watch them, think they’re great documentaries, and then spend the next hour talking about how upset it made them. And other people

watch them and then ask what's for dinner. You think that affects people's opinions on socialism much?

Imagine a proposal to institute a tax that would raise money for some effort to help mine workers in some way. Upon hearing of it, different people would have an emotional burst of pain of a certain size at the thought of hearing of a tax, and an emotional burst of pain of a different size at the thought of considering the mine workers. Neither of these bursts of pain would be proportional to the actual size of the problem as measured in some sort of ideal utilon currency (note especially scope insensitivity). But the brain very often makes decisions by comparing those two bursts of pain (see How We Decide or just the insula article above) and then comes up with reasons for the decision. So all the important issues like economic freedom and labor policy and maximizing utility and suchwhat get subordinated to whether you're secreting more neurotransmitters in response to money loss or images of sad coal miners.

If this theory were true, we would expect to find neurological differences in people of different political opinions. Ta da! A long list of neurological findings that differ in liberals and conservatives. Linking the startle reflex and the disgust reaction to the policies favored by these groups is left as a (very easy) exercise for the reader[3].

This may require some moderation of our political opinions on issues where we think we're far from the neurological norm. For example, I am no longer so confident that noise is such a big problem for everyone that we would all be better off if there were strict regulations on it. But I hope Bill will consider that some people may be so sensitive to noise that not everyone can just shrug it off, and so there may be a case for at least some regulation of it. Likewise, even though I don't mind

taxes too much, if my goal is a society where most people are happy I need to consider that a higher tax rate will decrease other people's happiness much more quickly than it decreases mine.

Other than that, it's just a general message of pessimism. If people's political opinions come partly from unchangeable anatomy, it makes the program of overcoming bias in politics a lot harder, and the possibility of coming up with arguments good enough to change someone else's opinion even more remote.

Footnotes

1) I am suitably ashamed of my appeal to pathos; my only defense is that it is entirely true, that I have only just finished moving, and that this post is hopefully a more appropriate response.

2) Actually, it's more complicated than this, because I agonize over spending money when shopping. I seem to use different thought processes for normal budgeting, and I expect there are many processes going on more complex than just high versus low aversion to money loss.

3) Possibly *too* easy. It's easy to go from that data to an explanation of why conservatives worry more about terrorism, but then why don't they also worry more about global warming?

II. Probabilism

Confidence Levels Inside and Outside an Argument

Related to: Infinite Certainty

Suppose the people at FiveThirtyEight have created a model to predict the results of an important election. After crunching poll data, area demographics, and all the usual things one crunches in such a situation, their model returns a greater than 999,999,999 in a billion chance that the incumbent wins the election. Suppose further that the results of this model are your only data and you know nothing else about the election. What is your confidence level that the incumbent wins the election?

Mine would be significantly less than 999,999,999 in a billion.

When an argument gives a probability of 999,999,999 in a billion for an event, then probably the majority of the probability of the event is no longer in "But that still leaves a one in a billion chance, right?". The majority of the probability is in "That argument is flawed". Even if you have no particular reason to believe the argument is flawed, the background chance of an argument being flawed is still greater than one in a billion.

More than one in a billion times a political scientist writes a model, ey will get completely confused and write something with no relation to reality. More than one in a billion times a programmer writes a program to crunch political statistics, there will be a bug that completely invalidates the results. More than one in a billion times a staffer at a website publishes the results of a political calculation online, ey will

accidentally switch which candidate goes with which chance of winning.

So one must distinguish between levels of confidence internal and external to a specific model or argument. Here the model's internal level of confidence is 999,999,999/billion. But my external level of confidence should be lower, even if the model is my only evidence, by an amount proportional to my trust in the model.

Is That Really True?

One might be tempted to respond "But there's an equal chance that the false model is too high, versus that it is too low." Maybe there was a bug in the computer program, but it prevented it from giving the incumbent's real chances of 999,999,999,999 out of a *trillion*.

The prior probability of a candidate winning an election is 50%[1]. We need information to push us away from this probability in either direction. To push significantly away from this probability, we need strong information. Any weakness in the information weakens its ability to push away from the prior. If there's a flaw in FiveThirtyEight's model, that takes us away from their probability of 999,999,999 in of a billion, and back closer to the prior probability of 50%

We can confirm this with a quick sanity check. Suppose we know nothing about the election (ie we still think it's 50-50) until an insane person reports a hallucination that an angel has declared the incumbent to have a 999,999,999/billion chance. We would not be tempted to accept this figure on the grounds that it is equally likely to be too high as too low.

A second objection covers situations such as a lottery. I would like to say the chance that Bob wins a lottery with one billion players is 1/1 billion. Do I have to adjust this upward to cover

the possibility that my model for how lotteries work is somehow flawed? No. Even if I am misunderstanding the lottery, I have not departed from my prior. Here, new information really does have an equal chance of going against Bob as of going in his favor. For example, the lottery may be fixed (meaning my original model of how to determine lottery winners is fatally flawed), but there is no greater reason to believe it is fixed in favor of Bob than anyone else.[2]

Spotted in the Wild

The recent Pascal's Mugging thread spawned a discussion of the Large Hadron Collider destroying the universe, which also got continued on an older LHC thread from a few years ago. Everyone involved agreed the chances of the LHC destroying the world were less than one in a million, but several people gave extraordinarily low chances based on cosmic ray collisions. The argument was that since cosmic rays have been performing particle collisions similar to the LHC's zillions of times per year, the chance that the LHC will destroy the world is either literally zero, or else a number related to the probability that there's some chance of a cosmic ray destroying the world so miniscule that it hasn't gotten actualized in zillions of cosmic ray collisions. Of the commenters mentioning this argument, one gave a probability of 1/3*10^22, another suggested 1/10^25, both of which may be good numbers for the internal confidence of this argument.

But the connection between this argument and the general LHC argument flows through statements like "collisions produced by cosmic rays will be exactly like those produced by the LHC", "our understanding of the properties of cosmic rays is largely correct", and "I'm not high on drugs right now, staring at a package of M&Ms and mistaking it for a really intelligent argument that bears on the LHC question", all of

which are probably more likely than 1/10^20. So instead of saying "the probability of an LHC apocalypse is now 1/10^20", say "I have an argument that has an internal probability of an LHC apocalypse as 1/10^20, which lowers my probability a bit depending on how much I trust that argument".

In fact, the argument has a potential flaw: according to Giddings and Mangano, the physicists officially tasked with investigating LHC risks, black holes from cosmic rays might have enough momentum to fly through Earth without harming it, and black holes from the LHC might not[3]. This was predictable: this was a simple argument in a complex area trying to prove a negative, and it would have been presumptous to believe with greater than 99% probability that it was flawless. If you can only give 99% probability to the argument being sound, then it can only reduce your probability in the conclusion by a factor of a hundred, not a factor of 10^20.

But it's hard for me to be properly outraged about this, since the LHC did not destroy the world. A better example might be the following, taken from an online discussion of creationism[4] and apparently based off of something by Fred Hoyle:

> In order for a single cell to live, all of the parts of the cell must be assembled before life starts. This involves 60,000 proteins that are assembled in roughly 100 different combinations. The probability that these complex groupings of proteins could have happened just by chance is extremely small. It is about 1 chance in 10 to the 4,478,296 power. The probability of a living cell being assembled just by chance is so small, that you may as well consider it to be impossible. This means that the

> probability that the living cell is created by an intelligent creator, that designed it, is extremely large. The probability that God created the living cell is 10 to the 4,478,296 power to 1.

Note that someone just gave a confidence level of 10^4478296 to one and was wrong. This is the sort of thing that should *never ever happen*. This is possibly the *most wrong anyone has ever been*.

It is hard to say in words exactly how wrong this is. Saying "This person would be willing to bet the entire world GDP for a thousand years if evolution were true against a one in one million chance of receiving a single penny if creationism were true" doesn't even begin to cover it: a mere 1/10^25 would suffice there. Saying "This person believes he could make one statement about an issue as difficult as the origin of cellular life per Planck interval, every Planck interval from the Big Bang to the present day, and not be wrong even once" only brings us to 1/10^61 or so. If the chance of getting Ganser's Syndrome, the extraordinarily rare psychiatric condition that manifests in a compulsion to say false statements, is one in a hundred million, and the world's top hundred thousand biologists all agree that evolution is true, then this person should preferentially believe it is more likely that all hundred thousand have simultaneously come down with Ganser's Syndrome than that they are doing good biology[5]

This creationist's flaw wasn't mathematical; the math probably does return that number. The flaw was confusing the internal probability (that complex life would form completely at random in a way that can be represented with this particular algorithm) with the external probability (that life could form

without God). He should have added a term representing the chance that his knockdown argument just didn't apply.

Finally, consider the question of whether you can assign 100% certainty to a mathematical theorem for which a proof exists. Eliezer has already examined this issue and come out against it (citing as an example this story of Peter de Blanc's). In fact, this is just the specific case of differentiating internal versus external probability when internal probability is equal to 100%. Now your probability that the theorem is false is entirely based on the probability that you've made some mistake.

The many mathematical proofs that were later overturned provide practical justification for this mindset.

This is not a fully general argument against giving very high levels of confidence: very complex situations and situations with many exclusive possible outcomes (like the lottery example) may still make it to the 1/10^20 level, albeit probably not the 1/10^4478296. But in other sorts of cases, giving a very high level of confidence requires a check that you're not confusing the probability inside one argument with the probability of the question as a whole.

Footnotes

1. Although technically we know we're talking about an incumbent, who typically has a much higher chance, around 90% in Congress.

2. A particularly devious objection might be "What if the lottery commissioner, in a fit of political correctness, decides that "everyone is a winner" and splits the jackpot a billion ways? If this would satisfy your criteria for "winning the lottery", then this mere possibility should indeed move your probability upward. In fact, since there is probably greater than

a one in one billion chance of this happening, the majority of your probability for Bob winning the lottery should concentrate here!"

3. Giddings and Mangano then go on to re-prove the original "won't cause an apocalypse" argument using a more complicated method involving white dwarf stars.

4. While searching creationist websites for the half-remembered argument I was looking for, I found what may be my new favorite quote: "Mathematicians generally agree that, statistically, any odds beyond 1 in 10 to the 50th have a zero probability of ever happening."

5. I'm a little worried that five years from now I'll see this quoted on some creationist website as an actual argument.

Schizophrenia and Geomagnetic Storms

Today I learned that some people hear voices because the stars are beaming invisible energy into their pineal glands.

I mention this because it sounds, not just crazy, but *textbook* crazy, the sort of thing a hack writer would make a crazy character after abandoning all subtlety. And it is always interesting when these sorts of things tun out to be not just not-crazy, but *true*, because it makes me wonder what else I am missing.

There's an association, well-supported but still not widely accepted, between schizophrenia and the geomagnetic storms caused by solar wind. The association may be mediated by the pineal gland, which forms differently in utero depending on the level of magnetic fluctuation in the environment. Here's one of the studies involved. So next time that wild-eyed bearded man on the street shouts that the rays shooting through his pineal gland are controlling his actions, just smile and tell him you already know.

Also, people born in February and March who grow up to be baseball players are most likely to be first basemen; people born in August or September are more likely to play at third. The explanation, as far as I understand it which is not very, has something to do with sunlight levels in the fourth week after conception altering levels of oxidizing chemicals in the mother's blood that affect the development of cerebral asymmetry in the fetus and affect the relative dominance of its right and left hands. It's all in Conception season and cerebral asymmetries among American baseball players.

And to think there are still people who say science is *boring*.

Talking Snakes: A Cautionary Tale

I particularly remember one scene from Bill Maher's "Religulous". I can't find the exact quote, but I will try to sum up his argument as best I remember.

> Christians believe that sin is caused by a talking snake. They may have billions of believers, thousands of years of tradition behind them, and a vast literature of apologetics justifying their faith - but when all is said and done, they're adults who believe in a talking snake.

I have read of the absurdity heuristic. I know that it is not carte blanche to go around rejecting beliefs that seem silly. But I was still sympathetic to the talking snake argument. After all…a *talking snake*?

I changed my mind in a Cairo cafe, talking to a young Muslim woman. I let it slip during the conversation that I was an atheist, and she seemed genuinely curious why. You've all probably been in such a situation, and you probably know how hard it is to choose just one reason, but I'd been reading about Biblical contradictions at the time and I mentioned the myriad errors and atrocities and contradictions in all the Holy Books.

Her response? "Oh, thank goodness it's that. I was afraid you were one of those crazies who believed that monkeys transformed into humans."

I admitted that um, well, maybe I sorta kinda might in fact believe that.

It is hard for me to describe exactly the look of shock on her face, but I have no doubt that her horror was genuine. I may have been the first flesh-and-blood evolutionist she ever met.

“But…” she looked at me as if I was an idiot. “Monkeys don’t change into humans. What on Earth makes you think monkeys can change into humans?”

I admitted that the whole process was rather complicated. I suggested that it wasn’t exactly a Optimus Prime-style transformation so much as a gradual change over eons and eons. I recommended a few books on evolution that might explain it better than I could.

She said that she respected me as a person but that quite frankly I could save my breath because there was no way any book could possibly convince her that monkeys have human babies or whatever sort of balderdash I was preaching. She accused me and other evolution believers of being too willing to accept absurdities, motivated by our atheism and our fear of the self-esteem hit we’d take by accepting Allah was greater than ourselves.

It is not clear to me that this woman did anything differently than Bill Maher. Both heard statements that sounded so crazy as to not even merit further argument. Both recognized that there was a large group of people who found these statements plausible and had written extensive literature justifying them. Both decided that the statements were so absurd as to not merit examining that literature more closely. Both came up with reasons why they could discount the large number of believers because those believers must be biased.

I post this as a cautionary tale as we discuss the logic or illogic of theism. I propose taking from it the following lessons:

- The absurdity heuristic doesn’t work very well.

- Even on things that sound really, really absurd.

- If a large number of intelligent people believe something, it deserves your attention. After you've studied it on its own terms, then you have a right to reject it. You could still be wrong, though.

- Even if you can think of a good reason why people might be biased towards the silly idea, thus explaining it away, your good reason may still be false.

- If someone cannot explain why something is not stupid to you over twenty minutes at a cafe, that doesn't mean it's stupid. It just means it's complicated, or they're not very good at explaining things.

- There is no royal road.

(special note to those prone to fundamental attribution errors: I do not accept theism. I think theism is wrong. I think it can be demonstrated to be wrong on logical grounds. I think the nonexistence of talking snakes is evidence against theism and can be worked into a general argument against theism. I just don't think it's as easy as saying "talking snakes are silly, therefore theism is false." And I find it embarrassing when atheists say things like that, and then get called on it by intelligent religious people.)

Arguments from My Opponent Believes Something

1. **Argument From My Opponent Believes Something, Which Is Kinda Like Believing It On Faith, Which Is Kinda Like Them Being A Religion**:
"The high priests of the economic orthodoxy take it on faith that anyone who doubts the market is a heretic who must be punished."

2. **Argument From My Opponent Believes Something, Which Means They Believe It Is The Answer To One Question, Which Is Kinda Like Believing It Is The Answer To All Questions, But It Isn't**: "Statists believe government can solve all our problems. They need to understand the world doesn't work that way."

3. **Argument From My Opponent Believes Something, Which Is Kinda Like Believing It Really Strongly, Which Is Kinda Like Being A Fanatic**: "Environmentalist extremists are fanatically obsessed over saving the planet, refusing to even consider any contradictory ideas."

4. **Argument From My Opponent Believes Something, Which Is Kinda Like Believing It Blindly With 100% Certainty**:
"Some people blindly trust science to always be correct about everything, but we need to remember that even scientists can make mistakes."

5. **Argument From My Opponent Believes Something, Which Is Kinda Like Having An Ideology, Which Means They Are Ideologues**: "Ideologies are false idols, attempts to replace thought with mindless obedience. And one such

ideology is the dogma of feminism. Therefore, we need to start being much more critical about feminism."

6. **Argument From My Opponent Believes Something, Which Is Kinda Like Hating The People Who Don't Believe In It, And Hatred Is Wrong**: "People need to get over their frothing hatred for euthanasia."

7. **Argument From My Opponent Believes Something, Which Is Kinda Like Saying That That One Belief Should Be The Sole Determinant Of Our Entire Aesthetic Sensibility**: "Sure, we could legalize contraception. But do we really want to enshrine the value that human fertility is evil, and that new human life is a 'failure' to be avoided?"

8. **Argument From My Opponent Believes Something, Which Might Suggest A Course Of Action, But A Suggestion Is Kinda Like An Obligation, And She Has No Right To Order Me Around**: "Some people want to liberalize immigration laws, but our country is under no obligation to let in any foreigner who asks."

9. **Argument From My Opponent Believes Something, Which Might Suggest A Course Of Action, Which Could In Theory Be Implemented Through Violence, And Violence Is Wrong**: "Transhumanists think AI may be dangerous, but this could encourage people to kill AI researchers, so holding this belief is irresponsible." Or, "Environmentalist condemnations of the oil industry encourage eco-terrorist attacks on oil workers."

10. **Argument From My Opponent Believes Something, Which Might Suggest A Course Of Action, And Suggestions Could In Theory Stigmatize People Who Don't Do Them**: "People say smoking is dangerous and unhealthy,

but this just serves to stigmatize smokers and make them feel unwelcome in society."

For best effect, combine all ten as densely as possible:

> It is an unchallengeable orthodoxy that you should wear a coat if it is cold out. Day after day we hear shrill warnings from the high priests of this new religion practically seething with hatred for anyone who might possibly dare to go out without a winter coat on. But these ideologues don't realize that just wearing more jackets can't solve all of our society's problems. Here's a reality check – no one is under any obligation to put on any clothing they don't want to, and North Face and REI are not entitled to your hard-earned money. All that these increasingly strident claims about jackets do is shame underprivileged people who can't afford jackets, suggesting them as legitimate targets for violence. In conclusion, do we really want to say that people should be judged by the clothes they wear? Or can we accept the unjacketed human body to be potentially just as beautiful as someone bundled beneath ten layers of coats?

EDIT: *I'm not claiming these aren't real problems, I'm claiming they're things that they are fully general arguments – you can accuse anyone of them and no one can ever prove you're wrong. For example, some things really are religions (Christianity, for example), but you can accuse any position of being "a religion" merely by virtue of it being a belief that people hold. Therefore, we should be extremely skeptical of arguments where "X is a religion" is doing the work.*

Statistical Literacy Among Doctors Now Lower Than Chance

Good news! 42% of doctors can correctly answer a true-false question on p-values! That's only 8% worse than a coin flip!

And this paragraph is your friendly reminder that six months after this study was published, the FDA decided it was unsafe for individuals to look at their own genome since they might misunderstand the risks involved. Instead, they must rely on their doctor. I am sure that statisticians and math professors making life-changing health or reproductive decisions feel perfectly confident being at the mercy of people whose statistics knowledge is worse than chance.

Now that I've got the sensationalism out of the way, let's look at this study more closely.

The sample is 4000 Ob/Gyn residents. Ob/Gyn is a prestigious specialty that's able to select people with very good grades in medical school, so we're not looking at dummies here. These residents (beginning doctors) did a bit worse than more experienced doctors (whose performance was still not stellar). I don't know whether this reflects doctors learning more about statistics as they progress, better statistical education in Ye Olde Days than in the current generation, or both.

The study looked at two questions. First was the one I mentioned above: "True or false: the p-value is the probability that the null hypothesis is correct". The correct answer is "false" – the p-value is the chance of obtaining results at least as extreme as those actually obtained if the null hypothesis were true. 42% correctly said it was false, 46% said it was true, and 12% didn't even want to hazard a guess.

The question seems sketchy to me. It is indeed technically false, but it seems pretty close to the truth. If I were asked to explain why the definition as given was false, the best I could do is say that your probability of the null hypothesis being true should take into account both something like your p-value, and your prior. But since no one ever receives Bayesian statistical education, I am not sure it is fair to expect a doctor to be able to generate that objection. What I would want a doctor to know is that the lower the p-value, the more conclusively the study has rejected the null hypothesis. The false definition as given accurately captures that key insight. So I'm not sure it proves anything other than doctors not being really nitpicky over definitions.

(which is also false, actually)

Next came very nearly the exact same question about mammogram results as Eliezer's Short Explanation Of Bayes Theorem. It offered five multiple-choice answers, so we would expect 20% correct by chance. Instead, 26% of doctors got it correct. What shocks me about this one is that the question very nearly does all the work for you and throws the right answer in your face. Compare the way it was phrased in Eliezer's example:

> 1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammographies. 9.6% of women without breast cancer will also get positive mammographies. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

to the way it was phrased on the obstetrician study:

> Ten out of every 1,000 women have breast cancer. Of these 10 women with breast cancer, 9 test positive. Of the 990 women without cancer, about 89 nevertheless test positive. A woman tests positive and wants to know whether she has breast cancer for sure, or at least what the chances are. What is the best answer?

The obstetrician study seems to be doing everything it can to guide people to the correct result, and 74% of people still got it wrong. And nitpicky definitions don't provide much of an excuse here.

There were three other results of this study worth highlighting.

First, people who got the statistics questions wrong were *more* likely to say they had good training in statistical literacy than those who did not, giving a rare demonstration of the Dunning-Kruger effect in the wild. Doctors who didn't know statistics were apparently so inadequate that they didn't realize there was any more to know, whereas those who did know some statistics at least had a faint inkling that something was missing.

Second, women rated their statistical literacy significantly worse than men did (note that a large majority of Ob/Gyn residents are women) but did not actually do any worse on the questions. This highlights an important limitation of self-report (tendency to confuse incompetence with humility) and probably has some broader gender-related implications as well.

And third, even though 42% of people got Question 1 correct and 26% of people Question 2, only 12% of people got both questions correct. Just from eyeballing those numbers, it doesn't look like getting one question right made you much

more likely to do better on the other. This is very consistent with most people lucking in to the correct answer.

I do not want to use this to attack doctors. Most doctors are technicians and not academics, and they cultivate, and should cultivate, only as much statistical knowledge as is useful for them. For a technician, “a p-value is that thing that gets lower when it means there’s really strong evidence” is probably enough. For a technician, “I can’t remember what exactly the positive predictive value of a mammogram is but it doesn’t matter because you should follow up all suspicious mammograms with further testing anyway” is probably enough.

But it really does seem relevant that only 12% of doctors can answer two simple statistics questions correctly when you’re trying to deny the entire non-doctor population access to certain information because only doctors are good enough at statistics to understand it.

Techniques for Probability Estimates

Utility maximization often requires determining a probability of a particular statement being true. But humans are not utility maximizers and often refuse to give precise numerical probabilities. Nevertheless, their actions reflect a "hidden" probability. For example, even someone who refused to give a precise probability for Barack Obama's re-election would probably jump at the chance to take a bet in which ey lost $5 if Obama wasn't re-elected but won $5 million if he was; such decisions demand that the decider covertly be working off of at least a vague probability.

When untrained people try to translate vague feelings like "It seems Obama will probably be re-elected" into a precise numerical probability, they commonly fall into certain traps and pitfalls that make their probability estimates inaccurate. Calling a probability estimate "inaccurate" causes philosophical problems, but these problems can be resolved by remembering that probability is "subjectively objective" - that although a mind "hosts" a probability estimate, that mind does not arbitrarily determine the estimate, but rather calculates it according to mathematical laws from available evidence. These calculations require too much computational power to use outside the simplest hypothetical examples, but they provide a standard by which to judge real probability estimates. They also suggest tests by which one can judge probabilities as well-calibrated or poorly-calibrated: for example, a person who constantly assigns 90% confidence to eir guesses but only guesses the right answer half the time is poorly calibrated. So calling a probability estimate "accurate" or "inaccurate" has a real philosophical grounding.

There exist several techniques that help people translate vague feelings of probability into more accurate numerical estimates. Most of them translate probabilities from forms without immediate consequences (which the brain supposedly processes for signaling purposes) to forms with immediate consequences (which the brain supposedly processes while focusing on those consequences).

Prepare for Revelation

What would you expect if you believed the answer to your question were about to be revealed to you?

In Belief in Belief, a man acts as if there is a dragon in his garage, but every time his neighbor comes up with an idea to test it, he has a reason why the test wouldn't work. If he imagined Omega (the superintelligence who is always right) offered to reveal the answer to him, he might realize he was expecting Omega to reveal the answer "No, there's no dragon". At the very least, he might realize he was worried that Omega would reveal this, and so re-think exactly how certain he was about the dragon issue.

This is a simple technique and has relatively few pitfalls.

Bet on it

At what odds would you be willing to bet on a proposition?

Suppose someone offers you a bet at even odds that Obama will be re-elected. Would you take it? What about two-to-one odds? Ten-to-one? In theory, the knowledge that money is at stake should make you consider the problem in "near mode" and maximize your chances of winning.

The problem with this method is that it only works when utility is linear with respect to money and you're not risk-

averse. In the simplest case I should be indifferent to a $100,000 bet at 50% odds that a fair coin would come up tails, but in fact I would refuse it; winning $100,000 would be moderately good, but losing $100,000 would put me deeply in debt and completely screw up my life. When these sorts of consideration become paramount, imagining wagers will tend to give inaccurate results.

Convert to a Frequency

How many situations would it take before you expected an event to occur?

Suppose you need to give a probability that the sun will rise tomorrow. “999,999 in a million” doesn’t immediately sound wrong; the sun seems likely to rise, and a million is a very high number. But if tomorrow is an average day, then your probability will be linked to the number of days it will take before you expect that the sun will fail to rise on at least one. A million days is three thousand years; the Earth has existed for far more than three thousand years without the sun failing to rise. Therefore, 999,999 in a million is too low a probability for this occurrence. If you think the sort of astronomical event that might prevent the sun from rising happens only once every three billion years, then you might consider a probability more like 999,999,999,999 in a trillion.

In addition to converting to a frequency across time, you can also convert to a frequency across places or people. What’s the probability that you will be murdered tomorrow? The best guess would be to check the murder rate for your area. What’s the probability there will be a major fire in your city this year? Check how many cities per year have major fires.

This method fails if your case is not typical: for example, if your city is on the losing side of a war against an enemy known to use fire-bombing, the probability of a fire there has nothing to do with the average probability across cities. And if you think the reason the sun might not rise is a supervillain building a high-tech sun-destroying machine, then consistent sunrises over the past three thousand years of low technology will provide little consolation.

A special case of the above failure is converting to frequency across time when considering an event that is known to take place at a certain distance from the present. For example, if today is April 10th, then the probability that we hold a Christmas celebration tomorrow is much lower than the 1/365 you get by checking on what percentage of days we celebrate Christmas. In the same way, although we know that the sun will fail to rise in a few billion years when it burns out its nuclear fuel, this shouldn't affect its chance of rising tomorrow.

Find a Reference Class

How often have similar statements been true?

What is the probability that the latest crisis in Korea escalates to a full-blown war? If there have been twenty crisis-level standoffs in the Korean peninsula in the past 60 years, and only one of them has resulted in a major war, then (war|crisis) = .05, so long as this crisis is equivalent to the twenty crises you're using as your reference class.

But finding the reference class is itself a hard problem. What is the probability Bigfoot exists? If one makes a reference class by saying that the yeti doesn't exist, the Loch Ness monster doesn't exist, and so on, then the Bigfoot partisan

might accuse you of assuming the conclusion - after all, the likelihood of these creatures existing is probably similar to and correlated with Bigfoot. The partisan might suggest asking how many creatures previously believed not to exist later turned out to exist - a list which includes real animals like the orangutan and platypus - but then one will have to debate whether to include creatures like dragons, orcs, and Pokemon on the list.

This works best when the reference class is more obvious, as in the Korea example.

Make Multiple Statements

How many statements could you make of about the same uncertainty as a given statement without being wrong once?

Suppose you believe France is larger than Italy. With what confidence should you believe it? If you made ten similar statements (Germany is larger than Austria, Britain is larger than Ireland, Spain is larger than Portugal, et cetera) how many times do you think you would be wrong? A hundred similar statements? If you think you'd be wrong only one time out of a hundred, you can give the statement 99% confidence.

This is the most controversial probability assessment technique; it tends to give lower levels of confidence than the others; for example, Eliezer wants to say there's a less than one in a million chance the LHC would destroy the world, but doubts he could make a million similar statements and only be wrong once. Komponisto thinks this is a failure of imagination: we imagine ourselves gradually growing tired and making mistakes, whereas this method only works if the accuracy of the millionth statement is exactly the same as the first.

In any case, the technique is only as good as the ability to judge which statements are equally difficult to a given statement. If I start saying things like “Russia is larger than Vatican City! Canada is larger than a speck of dust!” then I may get all the statements right, but it won’t mean much for my Italy-France example - and if I get bogged down in difficult questions like “Burundi is larger than Equatorial Guinea” then I might end up underconfident. In cases where there is an obvious comparison (“Bob didn’t cheat on his test”, “Sue didn’t cheat on her test”, “Alice didn’t cheat on her test”) this problem disappears somewhat.

Imagine Hypothetical Evidence

How would your probabilities adjust given new evidence?

Suppose one day all the religious people and all the atheists get tired of arguing and decide to settle the matter by experiment once and for all. The plan is to roll an n-sided numbered die and have the faithful of all religions pray for the die to land on “1”. The experiment will be done once, with great pomp and ceremony, and never repeated, lest the losers try for a better result. All the resources of the world’s skeptics and security forces will be deployed to prevent any tampering with the die, and we assume their success is guaranteed.

If the experimenters used a twenty-sided die, and the die comes up 1, would this convince you that God probably did it, or would you dismiss the result as a coincidence? What about a hundred-sided die? Million-sided? If a successful result on a hundred-sided die wouldn’t convince you, your probability of God’s existence must be less than one in a hundred; if a million-sided die would convince you, it must be more than one in a million.

This technique has also been denounced as inaccurate, on the grounds that our coincidence detectors are overactive and therefore in no state to be calibrating anything else. It would feel very hard to dismiss a successful result on a thousand-sided die, no matter how low the probability of God is. It might also be difficult to visualize a hypothetical where the experiment can't possibly be rigged, and it may be unfair to force subjects to imagine a hypothetical that would practically never happen (like the million-sided die landing on one in a world where God doesn't exist).

These techniques should be experimentally testable; any disagreement over which do or do not work (at least for a specific individual) can be resolved by going through a list of difficult questions, declaring confidence levels, and scoring the results with log odds. Steven's blog has some good sets of test questions (which I deliberately do *not* link here so as to not contaminate a possible pool of test subjects); if many people are interested in participating and there's a general consensus that an experiment would be useful, we can try to design one.

On First Looking into Chapman's "Pop Bayesianism"

I.

David Chapman keeps complaining that "Bayesianism" – as used to describe a philosophy rather than just a branch of statistics – is meaningless or irrelevant, yet is touted as being the Sacred Solution To Everything.

In my reply on his blog, I made the somewhat weak defense that it's not a disaster if a philosophy is not totally about its name. For example, the Baptists have done pretty well for themselves even though baptism is only a small part of their doctrine and indeed a part they share with lots of other denominations. The Quakers and Shakers are more than just people who move rhythmically sometimes, and no one gives *them* any grief about it.

But now I think this is overly pessimistic. I think Bayesianism is a genuine epistemology and that the only reason this isn't obvious is that it's a really *good* epistemology, so good that it's hard to remember that other people don't have it. So let me sketch two alternative epistemologies and then I'll define Bayesianism by contrast.

II.

Aristotelianism

Everyone likes to beat up on Aristotle, and I am no exception. An Aristotelian epistemology is one where statements are either true or false and you can usually figure out which by using deductive reasoning. Tell an Aristotelian a statement and, God help him, he will either agree or disagree.

Aristotelians are the sort of people who say things like "You can never really be an atheist, because you can't prove there's no God. If you were really honest you'd call yourself an agnostic." When an Aristotelian holds a belief, it's because he's damn well *proven* that belief, and if you say you have a belief but haven't proven it, you are a dirty cheater taking epistemic shortcuts.

Very occasionally someone will prove an Aristotelian wrong on one of his beliefs. This is shocking and traumatic, but it certainly doesn't mean that any of the Aristotelian's *other* beliefs might be wrong. After all, he's proven them with deductive reasoning. And deductive reasoning is 100% correct by definition! It's *logic*!

Anton-Wilsonism

Nobody likes to beat up on Robert Anton Wilson, and I consistently get complaints when I try. He and his ilk have *seen through* Aristotelianism. It's a sham to say you ever know things for certain, and there are a lot of dead white men who were cocksure about themselves and ended up being wrong. Therefore, the most virtuous possible epistemic state is to *not believe anything.*

This leads to nihilism, moral relativism, postmodernism, and mysticism. The truth cannot be spoken, because any assertion that gets spoken is just another dogma, and dogmas are the enemies of truth. Truth is in the process, or is a state of mind, or is [insert two hundred pages of mysterianist drivel that never really reaches a conclusion].

Bayesianism

"Epistemology X" is the synthesis of Aristotelianism and Anton-Wilsonism. It concedes that you are not certain of any of your beliefs. But it also concedes that you are not in a

position of global doubt, and that you can update your beliefs using evidence.

An Xist says things like "Given my current level of knowledge, I think it's 60% likely that God doesn't exist." If they encounter evidence for or against the existence of God, they might change that number to 50% or 70%. Or if they don't explicitly use numbers, they at least consider themselves to have strong leanings on difficult questions but with some remaining uncertainty. If they find themselves consistently over- or under-confident, they can adjust up or down until they reach either the certainty of Aristotelianism or the total Cartesian doubt of Anton-Wilsonism.

Epistemology X is both philosophically superior to its predecessors, in that it understands that you are neither completely omniscient nor completely nescient; instead, all knowledge is partial knowledge. And it is practically superior, in that it allows for the quantification of belief and therefore can have nice things like calibration testing and prediction markets.

What can we call this doctrine? In the old days it was known as probabilism, but this is unwieldy, and it refers to a variety practiced before we really understood what probability *was*. I think "Bayesianism" is an acceptable alternative, not just because Bayesian updating is the fundamental operation of this system, but because Bayesianism is the branch of probability that believes probabilities are degrees of mental credence and that allows for sensible probabilities of nonrepeated occurrences like "there is a God."

III.

"Jason" made nearly this exact same point on David's blog. David responds:

1) Do most people really think in black and white? Or is this a straw man?

2) Are numerical values a good way to think about uncertainty in general?

3) Does anyone actually consistently use numerical probabilities in everyday situations of uncertainty?

The discussion between David and Jason then goes off on a tangent, so let me give *my* answer to some of these questions.

Do people really think in black and white? Or in my formulation, is the "Aristotelian" worldview really as bad as all that? David acknowledges the whole "You can't really be an atheist because…" disaster, but says belief in God is a special case because of tribal affiliation.

I have consistently been tempted to agree with David – my conception of Aristotelianism certainly *sounds* like a straw man. But I think there are some inferential distances going on here. A year or so ago, my friend Ari wrote of Less Wrong:

> I think there's a few posts by Yudkowsky that I think deserve the highest praise one can give to a philosopher's writing: That, on rereading them, I have no idea what I found so mindblowing about them the first time. Everything they say seems patently obvious now!

Obviously not everyone gets this Bayesian worldview from Less Wrong, but I share this experience of "No, everything there is obvious, surely I must always have believed it" while having a vague feeling that there had been something extremely revolutionary-seeming to it at the time. And I have memories.

I remember how some of my first exposure to philosophy was arguing against Objectivists in my college's Objectivist Club. I remember how Objectivism absolutely lampshades Aristotelianism, how the head of the Objectivist Club tried very patiently to walk me through a deductive proof of why Objectivism was was correct from one of Rand's books. "It all starts with A = A," he told me. "From there, it's just logic." Although I did not agree with the proof itself, I don't remember finding anything objectionable in the methodology behind it, nor did any of the other dozen-odd people there.

I remember talking to my father about some form of alternative-but-not-implausible medicine. It might have been St. John's Wort – which has an evidence base now, but this was when I was very young. "Do you think it works?" I asked him. "There haven't been any studies on it," he said. "There's no evidence that it's effective." "Right," I said, "but there's quite a bit of anecdotal evidence in its favor." "But that's not proof," said my father. "You can't just start speculating on medicines when you don't have any proof that they work." Now, if I were in my father's shoes today, I might *still* make his same argument based on a more subtle evidence-based medicine philosophy, but the point was that at the time I felt like we were missing something important that I couldn't quite put my finger on, and looking back on the conversation, that thing we were missing is obviously the notion of probabilistic reasoning. From inside I know *I* was missing it, and when I asked my father about this a few years ago he completely failed to understand what relevance that could possibly have to the question, so I feel confident saying he was missing it too.

I remember hanging out with a group of people in college who all thought Robert Anton Wilson was the coolest thing since sliced bread, and it was *explicitly* because he said we didn't

have to believe things with certainty. I'm going to get the same flak I always get for this, but Robert Anton Wilson, despite his brilliance as a writer and person, has a *really dumb* philosophy. The *only* context in which it could possibly be attractive – and I say this as someone who went around quoting Robert Anton Wilson like nonstop for several months to a year – is if it was a necessary countermeasure to an *even worse* epistemology that we had been hearing our entire lives. What philosophy is this? Anton Wilson explicitly identifies it as the Aristotelian philosophy of deductive certainty.

And finally, I remember a rotation in medical school. I and a few other students were in a psychiatric hospital, discussing with a senior psychiatrist whether to involuntarily commit a man who had made some comments which sort of kind of sounded maybe suicidal. I took the opposing position: "In context, he's upset but clearly not at any immediate risk of killing himself." One of the other students took the opposite side: "If there's *any chance* he might shoot himself, it would be irresponsible to leave him untreated." This annoyed me. "There's "some chance" *you* might shoot yourself. Where do we draw the line?" The other student just laughed. "No, we're being serious here, and if you're not totally certain the guy is safe, he needs to be committed."

(before Vassar goes off on one of his "doctors are so stupid, they don't understand anything" rants, I should add that the senior psychiatrist then stopped the discussion, backed me up, and explained the basics of probability theory.)

So do most people really think in black and white? Ambiguous. I think people don't account for uncertainty in Far Mode, but do account for it in Near Mode. I think if you explicitly ask people "Should you take account of uncertainty?" they will say "yes", but if you ask them "Should

you commit anybody who has *any chance at all* of shooting themselves?" they will also say yes – and if you ask them "What chance of someone being a terrorist is too high before you let them fly on an airplane, and don't answer 'zero'?" they will look at you as if you just grew a second head.

In short, they are not actually idiots, but they have no coherent philosophical foundation for their non-idiocy, and this tends to show through at inconvenient times.

Probability theory in general, and Bayesianism in particular, provide a coherent philosophical foundation for not being an idiot.

Now in general, people don't need coherent philosophical foundations for anything they do. They don't need grammar to speak a language, they don't need classical physics to hit a baseball, and they don't need probability theory to make good decisions. This is why I find all the "But probability theory isn't that useful in everyday life!" complaining so vacuous.

"Everyday life" means "inside your comfort zone". You don't need theory inside your comfort zone, because you already navigate it effortlessly. But sometimes you find that the inside of your comfort zone isn't so comfortable after all (my go-to grammatical example is answering the phone "Scott? Yes, this is him.") Other times you want to leave your comfort zone, by for example speaking a foreign language or creating a conlang.

When David says that "You can't possibly be an atheist because…" doesn't count because it's an edge case, I respond that it's *exactly* the sort of thing that should count because it's people trying to actually *think* about an issue outside their comfort zone which they can't handle on intuition alone. It turns out when most people try this they fail miserably. If you are the sort of person who likes to deal with complicated

philosophical problems outside the comfortable area where you can rely on instinct – and politics, religion, philosophy, and charity all fall in that area – then it's *really nice* to have an epistemology that doesn't suck.

IV.

A while ago I wrote a post called Arguments From My Opponent Believes Something (read it!) which I feel was sorta misunderstood.

I made fun of people who attack arguments on the grounds that "this is like a religion!" or "some people say this solves everything!". Some people pointed out that, in fact, many things *are* like religions (religions being just the most obvious example) and sometimes people *do* fall into the trap of claiming their pet theory can explain everything.

And okay, I agree.

Those arguments aren't dangerous because they're never true. They're dangerous because you can always make them, whether they're true or not.

You can take *any* position in *any* argument and accuse the proponents of believing it fanatically. And then you're done. There's no good standard for fanaticism. Some people want to end the war in Afghanistan? Simply call them "anti-war fanatics". You don't have to prove anything, and even if the anti-war crowd object, they're now stuck objecting to the "fanatic" label rather than giving arguments against the war.

(if a candidate is stuck arguing "I'm not a child molester", then he has already lost the election, whether or not he manages to convince the electorate of his probable innocence)

And then when the war goes bad and hindsight bias tells us it was a terrible idea all along, you can just say "Yes, people like

me were happy to acknowledge the excellent arguments about the war. It was just *you guys* being *fanatics* about it all the time which *turned everyone else off*."

One of the wisest things I ever saw on Twitter (which is a low bar, sort of like "one of Hitler's most tolerant speeches") was on arrogance. "If someone you never met calls you 'arrogant', it means he can't find anything else," the tweet said. "Otherwise, he would have called you 'wrong'." My quotes file mysteriously labels this as "Heuristic 81 from Twitter", without giving a source or any hint on what the other eighty heuristics might be.

The Arguments From My Opponent Believes Something are a lot like accusations of arrogance. They're last-ditch attempts to muddy up the waters. If someone says a particular theory doesn't explain everything, or that it's elitist, or that it's being turned into a religion, that means they can't find anything else.

Otherwise they would have called it wrong.

Utilitarianism for Engineers

(title a reference to this SMBC comic)

I've said before that it's impossible to compare interpersonal utilities in theory but pretty easy in practice. Every time you give up your seat on the subway to an old woman with a cane, you're doing a quick little interpersonal utility calculation, and as far as I can tell you're getting it right.

The lack of the theory still grates, though, and I appreciate it whenever people come up with something halfway between theory and practice; some hack that lets people measure utilities rigorously enough to calculate surprising results, but not so rigorously that you run up against the limits of the math. The best example of this is the health care concept of QALYs, Quality Adjusted Life Years.

The Life Year part is pretty simple. If you only have $20,000 to spend on health care, and you can buy malaria drugs for $1,000 or cancer drugs for $10,000, what do you do? Suppose on average one out of every ten doses of malaria drugs save the life of a child who goes on to live another sixty years. And suppose on average every dose of cancer drug saves the life of one adult who goes on to live another twenty years.

In that case, each dose of malaria drug saves on average six life years, and each dose of cancer drug saves on average twenty life years. Given the cost of both drugs, your $20000 invested in malaria could save 120 life years, and your $20000 invested in cancer could save 40 life years. So spend the money on malaria (all numbers are made up, but spending health resources on malaria is usually a good decision).

The Quality Adjusted part is a little tougher. Suppose that the malaria drug also made everyone who used it break out in hideous blue boils, but the cancer drug made them perfectly healthy in every way. We would want to penalize the malaria drug for this. How much do we penalize it? Some amount based on how much people disvalue hideous blue boils versus being perfectly healthy versus dying of malaria. A classic question is "If you were covered in hideous blue boils, and there were a drug that had an X% chance of making you perfectly healthy but a (100 - x%) chance of killing you, would you take it?" And if people on average say yes when X = 50, then we may value a life-year spend with hideous blue boils at only 50% that of a life year spent perfectly healthy.

So now instead of being 120 LY from malaria versus 40 LY from cancer, it's 60 LY from malaria versus 40 from cancer; we should still spend the money on the malaria drug, but it's not quite as big a win any more.

[I have gone back and edited parts this post three times, and each time I read that last sentence, I think of a spaceship a hundred twenty light years away from the nearest malaria parasite.]

Some public policy experts actually use utilitarian calculations over QALYs to make policy. I read an excellent analysis once by some surgeons arguing which of two treatment regimens for colon cancer was better. One treatment regimen included much stronger medicine that had much worse side effects. The surgeon supporting it laboriously went through the studies showing increased survival rates, subtracted out QALYs for years spent without a functional colon, found the percent occurrence of each side effect and subtracted out QALYs based on its severity, and found that on average the stronger

medicine gained patients more utility than the weaker medicine - let's say 0.5 extra QALYs.

Then he compared the cost of the medicine to the cost of other interventions that on average produced 0.5 extra QALYs. He found that his medicine was more cost-effective than many other health care interventions that returned the same benefits, and therefore recommended it both to patients and insurance bureaucrats.

As far as I can tell, prescribing that one colon cancer medicine is now on sounder epistemological footing than any other decision any human being has ever made.

Towards A More General Hand-Wavy Pseudotheory

So if we can create a serviceable hack that lets us sort of calculate utility in medicine, why can't we do it for everything else?

I'm not saying QALYs are great. In fact, when other people tried the colon cancer calculation they got different results by about an order of magnitude.

But a lot of our social problems seem to be things where the two sides differ by at least an order of magnitude - I don't think even the most conservative mathematician could figure out a plausible way to make the utilitarian costs of gay marriage appear to exceed the benefits. Even a biased calculation would improve political debate: people would be forced to say which term in the equation was wrong, instead of talking about how the senator proposing it had an affair or something. And it could in theory provide the same kind of imperfect-but-useful-for-coordination focal point as a prediction market.

Okay, sorry. I'm done trying to claim this is a useful endeavor. I just think it would be really fun to try. If I need to use the excuse that I'm doing it for a constructed culture in a fictional setting I'm designing, I can pull that one out too (it is in fact true). So how would one create a general measure as useful as the QALY?

Start with a bag of five items, all intended to be good in some way very different from that in which the others are good:

1. $10,000 right now.
2. +5 IQ points
3. Sex with Scarlet Johannson
4. Saving the Amazon rainforest
5. Landing a man on Mars

A good hand-wavy pseudotheory of utility would have to be able to value all five of these goods in a common currency, and by extension relative to one another. We imagine asking several hundred people a certain question, and averaging their results. In some cases the results would be wildly divergent (for example, values of 3 would differ based on sex and sexual orientation) but they might still work as a guide, in the same way that believing each person to have one breast and one testicle would still allow correct calculation of the total number of breasts and testicles in society.

Let's start with the most impossible problem first: what question would we be asking people and then averaging the results of?

The VNM axioms come with a built in procedure for *part* of this - a tradeoff of probabilities. Would you rather save the Amazon rainforest, or have humankind pull off a successful Mars mission? If you prefer saving the rainforest, your next question is: would you rather have a 50% chance of saving the

rainforest, or a 100% chance of a successful Mars mission? If you're indifferent between the second two, we can say that saving the Amazon is worth twice as many utils as a Mars mission for you. If you'd also be indifferent between a 50% chance at a Mars mission and a 100% chance of $10,000, then we can say that - at least within those three things - the money is worth 1 util, the Mars landing is worth 2 utils, and the rainforest is worth 4 utils.

The biggest problem here is that - as has been remarked *ad nauseum* - this is only ordinal rather than cardinal and so makes interpersonal utility comparisons impossible. It may be that I have stronger desires than you on everything, and this method wouldn't address that. What can we turn into a utility currency that can be compared across different people?

The economy uses money here, and it seems to be doing pretty well for itself. But the whole point of this exercise is to see if we can do better, and money leaves much to be desired. Most important, it weights people's utility in proportion to how much money they have. A poor person who really desperately wants a certain item will be outbid by a rich person who merely has a slight preference for it. This produces various inefficiences (if you can call, for example, a global famine killing millions an "inefficiency") and is exactly the sort of thing we want a hand-wavy pseudotheory of utility to be able to outdo.

We could give everyone 100 Utility Points, no more, no less, and allow these to be used as currency in exactly the same way the modern economy uses money as currency. But is utility a zero sum game within agents? Suppose I want a plasma TV. Then I get cancer. Now I really really want medical treatment. Is there some finite amount of wanting that my desire for

cancer treatment takes up, such that I want a plasma TV less than I did before? I'm not sure.

Just as you can assign logarithmic scoring rules to beliefs to force people to make them correspond to probabilities, maybe you can assign them to wants as well? So we could ask people to assign 100% among the five goods in our basket, with the percent equalling the probability that each event will happen, and use some scoring rule to prevent people from assigning all probability to the event they want the most? Mathematicians, back me up on this?

The problem here is that there's no intuitive feel for it. We'd just be assigning numbers. Just as probability calibration is bad, I bet utility calibration is also bad. Also, comparing things specific to me (like me getting $10,000) plus things general to the world (saving the rainforest) is hard.

What about just copying the QALY metric completely? How many years (days?) of life would you give up for a free $10,000? How about to save the rainforest? This one has the advantage of being easy-to-understand and being a real choice that someone could ponder on. And since most people have similar expected lifespans, it's more directly comparable than money.

But this too has its problems. I visualize the last few years of my life being spent in a nursing home - I would give those up pretty easily. The next few decades are iffy. And it would take a lot to make me take forty years off my life, since that would bring my death very close to the present. On the other hand, some things I want *more* than this scale could represent; if I would gladly give my own life to solve poverty in Africa, how many QALYs is that? The infinity I would be willing to give, or the fifty or so I've actually got. If we limit me to fifty, that

suggests I place the same value on solving poverty in Africa as on solving poverty all over the world, which is just dumb.

Someone in the Boston Less Wrong meetup group yesterday suggested pain. How many seconds of some specific torture would you be willing to undergo in order to gain each good? This has the advantage of being testable: we can for example offer a randomly selected sample of people the opportunity to actually undergo torture in order to get $10,000 or whatever in order to calibrate their assessments ("Excuse me, Ms. Johansson, would you like to help us determine people's utility functions?")

But pain probably scales nonlinearly, different tortures are probably more or less painful to different people, and as I mentioned the last time this was brought up society would get taken over by a few people with Congenital Insensitivity To Pain Disorder.

Maybe the best option would be simple VNM comparisons with a few fixed interpersonal comparison points that we expect to be broadly the same among people. A QALY would be one. A certain amount of pain might be another. If we were really clever, we could come up with a curve representing the utility of money at different wealth levels, and use the utility of money transformed via that curve as a third.

Then we just scale everyone's curve so that the comparison points are as close to other people's comparison points as possible, stick it on the interval between one and zero, and call that a cardinal utility function.

Among the horrible problems that would immediately ruin everything are:

- massive irresolvable individual differences (like the sexual orientation thing, or value of money at different wealth levels)

- people exaggerating in order to inflate the value of their preferred policies, difficulty specifying the situation (what exactly needs to occur for the Amazon to be considered "saved"?)

- separating base-level preferences from higher-level preferences (do you have a base level preference against racism, or is your base level preference for people living satisfactory lives and you think racism makes people's lives worse; if the latter we risk double-counting against racism)

- people who just have stupid preferences not based on smart higher-level preferences (THE ONLY THING I CARE ABOUT IS GAY PEOPLE NOT MARRYING!!!)

- scaling the ends of the function (if I have a perfectly normal function but then put "making me supreme ruler of Earth" as 1000000000000000000x more important than everything else, how do we prevent that from making it into the results without denying that some people may really have things they value very very highly?)

- a sneaking suspicion that the scaling process might not be as mathematically easy as I, knowing nothing about mathematics, assume it ought to be.

I'd be very interested if anyone has better ideas along these lines, or stabs at solutions to any of the above problems. I'm not going to commit to actually *designing* a system like this, but it's been on my list of things to do if I ever get a full month semi-free, and if I can finish *Dungeons and Discourse* in time I might find myself in that position.

If It's Worth Doing, It's Worth Doing with Made-Up Statistics

I do not believe that the utility weights I worked on last week – the ones that say living in North Korea is 37% as good as living in the First World – are objectively correct or correspond to any sort of natural category. So why do I find them so interesting?

A few weeks ago I got to go to a free CFAR tutorial (you can hear about these kinds of things by signing up for their newsletter). During this particular tutorial, Julia tried to explain Bayes' Theorem to some, er, rationality virgins. I record a heavily-edited-to-avoid-recognizable-details memory of the conversation below:

Julia: So let's try an example. Suppose there's a five percent chance per month your computer breaks down. In that case…
Student: Whoa. Hold on here. That's not the chance my computer will break down.
Julia: No? Well, what do you think the chance is?
Student: Who knows? It might happen, or it might not.
Julia: Right, but can you turn that into a number?
Student: No. I have no idea whether my computer will break. I'd be making the number up.
Julia: Well, in a sense, yes. But you'd be communicating some information. A 1% chance your computer will break down is very different from a 99% chance.
Student: I don't know the future. Why do you want to me to pretend I do?
Julia: *(who is heroically nice and patient)* Okay, let's back up. Suppose you buy a sandwich. Is the sandwich probably

poisoned, or probably not poisoned?
Student: Exactly which sandwich are we talking about here?

In the context of a lesson on probability, this is a problem I think most people would be able to avoid. But the student's attitude, the one that rejects hokey quantification of things we don't actually know how to quantify, is a pretty common one. And it informs a lot of the objections to utilitarianism – the problem of quantifying exactly how bad North Korea shares some of the pitfalls of quantifying exactly how likely your computer is to break (for example, "we are kind of making this number up" is a pitfall).

The explanation that Julia and I tried to give the other student was that imperfect information still beats zero information. Even if the number "five percent" was made up (suppose that this is a new kind of computer being used in a new way that cannot be easily compared to longevity data for previous computers) it encodes our knowledge that computers are unlikely to break in any given month. Even if we are wrong by a very large amount (let's say we're off by a factor of four and the real number is 20%), if the insight we encoded into the number is sane we're still doing better than giving no information at all (maybe model this as a random number generator which chooses anything from 0 – 100?)

This is part of why I respect utilitarianism. Sure, the actual badness of North Korea may not be exactly 37%. But it's probably not twice as good as living in the First World. Or even 90% as good. But it's probably not two hundred times worse than death either. There is definitely nonzero information transfer going on here.

But the typical opponents of utilitarianism have a much stronger point than the guy at the CFAR class. They're not

arguing that utilitarianism fails to outperform zero information, they're arguing that it fails to outperform our natural intuitive ways of looking at things, the one where you just think "North Korea? Sounds awful. The people there deserve our sympathy."

Remember the Bayes mammogram problem? The correct answer is 7.8%; most doctors (and others) intuitively feel like the answer should be about 80%. So doctors – who are specifically trained in having good intuitive judgment about diseases – are wrong by an order of magnitude. And it "only" being *one* order of magnitude is not to the doctors' credit: by changing the numbers in the problem we can make doctors' answers as wrong as we want.

So the doctors probably would be better off explicitly doing the Bayesian calculation. But suppose some doctor's internet is down (you have NO IDEA how much doctors secretly rely on the Internet) and she can't remember the prevalence of breast cancer. If the doctor thinks her guess will be off by less than an order of magnitude, then making up a number and plugging it into Bayes will be more accurate than just using a gut feeling about how likely the test is to work. Even making up numbers based on basic knowledge like "Most women do not have breast cancer at any given time" might be enough to make Bayes Theorem outperform intuitive decision-making in many cases.

And a *lot* of intuitive decisions are off by way more than the make-up-numbers ability is likely to be off by. Remember that scope insensitivity experiment where people were willing to spend about the same amount of money to save 2,000 birds as 200,000 birds? And the experiment where people are willing to work harder to save one impoverished child than fifty impoverished children? And the one where judges give

criminals several times more severe punishments on average just before they eat lunch than just after they eat lunch?

And it's not just neutral biases. We've all seen people who approve wars under Republican presidents but are *horrified* by the injustice and atrocity of wars under Democratic presidents, even if it's just the same war that carried over to a different administration. If we forced them to stick a number on the amount of suffering caused by war before they knew what the question was going to be, that's a bit harder.

Thus is it written: "It's easy to lie with statistics, but it's easier to lie without them."

Some things work okay on System 1 reasoning. Other things work badly. Really really badly. Factor of a hundred badly, if you count the bird experiment.

It's hard to make a mistake in calculating the utility of living in North Korea that's off by a factor of *a hundred*. It's hard to come up with values that make a war suddenly become okay/abominable when the President changes parties.

Even if your data is completely made up, the way the 5% chance of breaking your computer was made up, the fact that you can apply normal non-made-up arithmetic to these made-up numbers will mean that you will very often *still* be less wrong than if you had used your considered and thoughtful and phronetic opinion.

On the other hand, it's pretty easy to accidentally Pascal's Mug yourself into giving everything you own to a crazy cult, which System 1 is good at avoiding. So it's nice to have data from both systems.

In cases where we really don't know what we're doing, like utilitarianism, one can still make System 1 decisions, but

making them with the System 2 data in front of you can change your mind. Like "Yes, do whatever you want here, just be aware that X causes two thousand people to die and Y causes twenty people an amount of pain which, in experiments, was rated about as bad as a stubbed toe".

And cases where we don't really know what we're doing have a wonderful habit of developing into cases where we *do* know what we're doing. Like in medicine, people started out with "doctors' clinical judgment obviously trumps everything, but just in case some doctors forgot to order clinical judgment, let's make some toy algorithms". And then people got better and better at crunching numbers and now there are cases where doctors should never use their clinical judgment under any circumstances. I can't find the article right now, but there are even cases where doctors armed with clinical algorithms consistently do worse than clinical algorithms without doctors. So it looks like at some point the diagnostic algorithm people figured out what they were doing.

I generally support applying made-up models to pretty much any problem possible, just to notice where our intuitions are going wrong and to get a second opinion from a process that has no common sense but is also lacks systematic bias (or else has unpredictable, different systematic bias).

This is why I'm disappointed that no one has ever tried expanding the QALY concept to things outside health care before. It's not that I think it will work. It's that I think it will fail to work in a different way than our naive opinions fail to work, and we might learn something from it.

EDIT: Edited to include some examples from the comments. I also really like ciphergoth's quote: "Sometimes pulling numbers out of your arse and using

them to make a decision is better than pulling a decision out of your arse."

Marijuana: Much More Than You Wanted to Know

This month I work on my hospital's Substance Abuse Team, which means we treat people who have been hospitalized for alcohol or drug-related problems and then gingerly suggest that maybe they should use drugs a little less.

The two doctors leading the team are both very experienced and have kind of seen it all, so it's interesting to get a perspective on drug issues from people on the front line. In particular, one of my attendings is an Obama-loving long-haired hippie who nevertheless vehemently opposes medical marijuana or any relaxation on marijuana's status at all. He says that "just because I'm a Democrat doesn't mean I have to support stupid policies I know are wrong" and he's able to back up his opinion with an impressive variety of studies.

To be honest, I had kind of forgotten that the Universe was allowed to contain negative consequences for legalizing drugs. What with all the mental energy it took protesting the the Drug War and getting outraged at police brutality and celebrating Colorado's recently permitting recreational cannabis use and so on, it had completely slipped my mind that the legalization of marijuana might have negative consequences and that I couldn't reject it out of hand until I had done some research.

So I've been doing the research. Not to try to convince my attending of anything – as the old saying goes, do not meddle in the affairs of attendings, because you are crunchy and taste good with ketchup – but just to figure out where exactly things stand.

I. Would Relaxation Of Penalties On Marijuana Increase Marijuana Use?

Starting in the 1970s, several states decriminalized possession of marijuana – that is, possession could not be penalized by jail time. It could still be penalized by fines and other smaller penalties, and manufacture and sale could still be punished by jail time.

Starting in the 1990s, several states legalized medical marijuana. People with medical marijuana cards, which in many cases were laughably easy to get with or without good evidence of disease, were allowed to grow and use marijuana, despite concerns that some of this would end up on the illegal market.

Starting last week, Colorado legalized recreational use of marijuana, as well as cultivation and sale (subject to heavy regulations). Washington will follow later this year, and other states will be placing measures on their ballots to do the same.

One should be able to evaluate to what degree marijuana use rose after these policy changes, and indeed, many people have tried – with greater or lesser levels of statistical sophistication.

The *worst* arguments in favor of this proposition are those like this CADCA paper, which note that states with more liberal marijuana laws have higher rates of marijuana use among teenagers than states that do not. The proper counterspell to such nonsense is *Reverse Causal Arrows* – could it not be that states with more marijuana users are more likely to pass proposals liberalizing marijuana laws? Yes it could. Even more likely, some third variable – let's call it "hippie attitudes" – could be behind both high rates of marijuana use and support for liberal marijuana regimes. The states involved are places like Colorado, California, Washington, and Oregon. I think

that speaks for itself. In case it doesn't, someone went through the statistics and found that these states had the highest rates of marijuana use among teens since *well* before they relaxed drug-related punishments. Argument successfully debunked.

A slightly more sophisticated version – used by the DEA here – takes the teenage marijuana use in a state one year before legalization of medical marijuana and compares it to the teenage marijuana use in a state one (or several years) after such legalization. They often find that it has increased, and blame the increase on the new laws. For example, 28% of Californians used marijuana before it was decriminalized in the 70s, compared to 35% a few years after. This falls victim to a different confounder – marijuana use has undergone some very large swings nationwide, so the rate of increase in medical marijuana states may be the same as the rate anywhere else. Indeed, this is what was going on in California – its marijuana use actually rose slightly *less* than the national average.

What we want is a study that compares the average marijuana use in a set of states before liberalization to the average marijuana use in the country as a whole, and then does the same after liberalization to see if the ratio has increased. There are several studies that purport to try this, of which by far the best is Johnston, O'Malley & Bachman 1981, which monitored the effect of the decriminalization campaigns of the 70s. They survey thousand of high school seniors on marijuana use in seven states that decriminalize marijuana both before and for five years after the decriminalization, and find absolutely no sign of increased marijuana use (in fact, there is a negative trend). Several other studies (eg Thies & Register 1993) confirm this finding.

There is only a hint of some different results. Saffer and Chaloukpa 1999 and Chaloupka, Grossman & Tauras 1999 try to use complicated econometric simulations to estimate the way marijuana demand will respond to different variables. They simulate (as opposed to detecting in real evidence) that marijuana decriminalization should raise past-year use by about 5 – 8%, but have no effect on more frequent use (ie a few more people try it but do not become regular users). More impressively, Model 1993 (a source of some exasperation for me earlier) finds that after decriminalization, marijuana-related emergency room visits went up (trying to interpret their tables, I think they went up by a whopping 90%, but I'm not sure of this). This is sufficiently different from every other study that I don't give it much weight, although we'll return to it later.

Overall I think the evidence is pretty strong that decriminalization probably led to no increase in marijuana use among teens, and may at most have led to a small single-digit increase.

Proponents of stricter marijuana penalties say the experiment isn't fair. In practice, decriminalization does not affect the average user very much – even in states without decriminalization, marijuana possession very rarely leads to jail time. The only hard number I have is from Australia, where in "non-decriminalized" Australian states only 0.3% of marijuana arrests lead to jail time, but a quick back-of-the-envelope calculation suggests US numbers are very similar. And even in supposedly decriminalized states, it's not hard for a cop who wants to get a pot user in jail to find a way (possession of even small amounts can be "possession with intent to sell" if someone doesn't like you). So the overall real difference between decriminalized and not decriminalized is small and it's not surprising the results are small as well. I

mostly agree with them; decriminalization is fine as far as it goes, but it's a bigger psychological step than an actual one.

The next major milestone in cannabis history was the legalization of medical marijuana. Anderson, Hansen & Rees (2012) did the same kind of study we have seen above, and despite trying multiple different measures of youth marijuana use found pretty much no evidence that medical marijuana legalization caused it to increase. Other studies find pretty much the same.

This could potentially suffer from the same problems as decriminalization studies – the laws don't always change the facts on the ground. Indeed, for about ten years after medical marijuana legalization, the federal government kept on prosecuting marijuana users even when their use accorded with state laws, and many states had so few dispensaries that in reality not a whole lot of medical marijuana was being given out. I haven't found any great studies that purport to overcome these problems.

When we examined decriminalization, we found that the studies based on surveys of teens looked pretty good, but that the one study that examined outcomes – marijuana-related ER visits – was a lot less encouraging. We find the same pattern here, and the rain on our parade is Chu 2013, who finds that medical marijuana laws increased marijuana-related arrests by 15-20% and marijuana-related drug rehab admissions by 10-15%.

So what's going on here? I have two theories. First, maybe medical marijuana use (and decriminalization) increase use among adults only. This could be because the system is working – giving adults access to medical marijuana while keeping it out of the hands of children – or because kids are

dumb and don't understand consequences but adults are more responsive to incentives and punishments. Second, we know that medical marijuana has twice as much THC as street marijuana. Maybe everyone keeps using the same amount of marijuana, but when medical marijuana inevitably gets diverted to the street, addicts can't handle it and end up behaving much worse than they expected.

Or the studies are wrong. Studies being wrong is always a pretty good bet.

I can't close this section without mentioning the Colorado expulsion controversy. Nearly everyone who teaches in Colorado says there has been an explosion of marijuana-related problems since medical marijuana was legalized. Meanwhile, the actual surveys of Colorado high school students say that marijuana use, if anything, is going down. A Colorado drug warrior has some strong objections to the survey results, but they center around not really being able to prove that there is a real downward trend (which is an entirely correct complaint) without denying that in fact they show no evidence at all of going *up*.

The consensus on medical marijuana seems to be that it does not increase teen marijuana use either, although there is some murky and suggestive evidence that it might increase illicit or dangerous marijuana use among adults.

There is less information on the effects of full legalization of marijuana, which has never been tried before in the United States. To make even wild guesses we will have to look at a few foreign countries plus some econometric simulations.

No one will be surprised to hear that the first foreign country involved is the Netherlands, which was famously permissive of cannabis up until a crackdown a few years ago. Despite

popular belief they never fully legalized the drug and they were still pretty harsh on production and manufacture; distribution, on the other hand, could occur semi-openly in coffee shops. This is another case where we have to be careful to distinguish legal regimes from actual effects, but during the period when there were actually a lot of pot-serving coffee shops, the Netherlands did experience an otherwise-inexplicable 35% rise in marijuana consumption relative to the rest of Europe. This is true even among teenagers, and covers both heavy use as well as occasional experimentation. Some scientists studying the Netherlands' example expect Colorado to see a similar rise; others think it will be even larger because the legalization is complete rather than partial.

The second foreign country involved is Portugal, which was maybe more of a decriminalization than a legalization case but which is forever linked with the idea of lax drug regimes in the minds of most Americans. They decriminalized all drugs (including heroin and cocaine) in 2001, choosing to replace punishment with increased treatment opportunities, and as we all have been told, no one in Portugal ever used drugs ever again, or even remembers that drugs exist. Except it turns out it's more complicated; for example, the percent of Portuguese who admit to lifetime use of drugs has doubled since the law took effect. Two very patient scientists have sifted through all the conflicting claims and found that in reality, the number of people who briefly experiment with drugs has gone way up, but the number of addicts hasn't, nor has the number of bad outcomes like overdose-related deaths. There are many more people receiving drug treatment, but that might just be because Portugal upped its drug treatment game in a separate law at the same time they decriminalized drugs. Overall they seem to have been a modest success – neither really raising nor

decreasing the number of addicts – but they seem more related to decriminalization (which we've already determined doesn't have much effect) than to legalization per se.

Returning to America, what if you just *ask* people whether they would use more marijuana if it's legal? Coloradans were asked if they plan to smoke marijuana once it becomes legal; comparing survey results to current usage numbers suggests 40% more users above the age of 18; it is unclear what the effect will be on younger teens and children.

Finally, we let the economists have their say. They crunch all the data and predict an increase of 50 – 100% based solely on the likely price drop (even with taxes factored in). And if there's one group we can trust to make infallible predictions about the future, it's economists.

Overall I find the Dutch evidence most convincing, and predict a 25 – 50% increase in adult marijuana use with legalization. I would expect a lower increase – 15 – 30% – among youth, but the data are also perfectly consistent with no increase at all.

Conclusion for this section: that decriminalization and legalization of medical marijuana do not increase youth marijuana use rates, although there is some shaky and indirect evidence they do increase adult use and bad behavior. There is no good data yet on full legalization, but there's good reason to think it would substantially increase adult use and it might also increase youth use somewhat.

II. Is Marijuana Bad For You?

About 9% of marijuana users eventually become addicted to the drug, exposing them to various potential side effects.

Marijuana smoke contains a lot of the same chemicals in tobacco smoke and so it would not be at all surprising if it had some of the same ill effects, like cardiovascular disease and lung cancer. But when people look for these effects, they can't find any increase in mortality among marijuana smokers. I predict that larger studies will one day pick something up, but for now let's take this at face value.

Much more concerning are the attempts to link marijuana to cognitive and psychiatric side effects. Meier et al (2012) analyzed a study of a thousand people in New Zealand and found that heavy marijuana use was linked to an IQ decline of 8 points. Rogeberg 2012 developed an alternative explanation – poor people saw their IQs drop in their 20s more than rich people because their IQs had been artificially inflated by schooling; what Meier et al had thought to be an effect of cannabis was really an effect of poor people having an apparent IQ drop and using cannabis more often. Meier et al pointed out that actually, poor people didn't use cannabis any more often than anyone else and effects remained when controlled for class. Other studies, like Fried et al (2002) find the same effect, and there is a plausible biological mechanism (cannabinoids something something neurotransmitters something brain maturation). As far as I can tell the finding still seems legit, and marijuana use does decrease IQ. It is still unclear whether this only applies in teenagers (who are undergoing a "sensitive period of brain development") or full stop.

More serious still is the link with psychosis. A number of studies have found that marijuana use is heavily correlated with development of schizophrenia and related psychotic disorders later in life. Some of them find relative risks as high as 2 – heavy marijuana use doubles your chance of getting

schizophrenia, which is already a moderately high 1%. But of course correlation is not causation, and many people have come up with alternative theories. For example, maybe people who are already kind of psychotic use marijuana to self-medicate, or just make poor life choices like starting drugs. Maybe people of low socioeconomic status who come from broken homes are more likely to both use marijuana and get schizophrenia. Maybe some gene both makes marijuana really pleasant and increases schizophrenia risk.

I know of three good studies attempting to tease out causation. Arseneault et al (2004) checks to see which came first – the marijuana use or the psychotic symptoms – and finds it was the marijuana use, thus supporting an increase in risk from the drug. Griffith-Lendering et al (2012) try the same, and find *bidirectional* causation – previous marijuana use seems to predict future psychosis, but previous psychosis seems to predict future marijuana use. A very new study from last month boxes clever and checks whether your marijuana use can predict schizophrenia in *your relatives*, and find that it does – presumably suggesting that genetic tendencies towards schizophrenia cause marijuana use and not vice versa (although Ozy points out to meet that the relatives of marijuana users are more likely to use marijuana themselves; the plot thickens). When a meta-analysis tries to control for all of these factors, they get a relative risk of 1.4 (they call it an odds ratio, but from their discussion section I think they mean relative risk).

Is this true, or just the confounders they failed to pick up? One argument for the latter is that marijuana use has increased very much over the past 50 years. If marijuana use caused schizophrenia, we would expect to see much more schizophrenia, but in fact as far as anyone can tell (which is

not very far) schizophrenia incidence is decreasing. The decrease might be due (maybe! if it even exists at all!) to obstetric advances which prevent fetal brain damage which could later lead to the disease. The effect of this variable is insufficiently known to pretend we can tease out some supposed contrary effect of increased marijuana use. Also, some people say that schizophrenia is increasing in young people, so who knows?

The *exact* nature of the marijuana-psychosis link is still very controversial. Some people say that marijuana causes psychosis. Other people say it "activates latent psychosis", a term without a very good meaning but which might mean that it pushes people on the borderline of psychosis – eg those with a strong family history but who might otherwise have escaped – over the edge. Still others say all it does is get people who would have developed psychosis eventually to develop it a few years earlier. You can read a comparison of all the different hypotheses here.

I've saved the most annoying for last: is marijuana a "gateway drug"? Would legalizing it make it more or less of a "gateway drug"? This claim seems tailor-made to torture statisticians. We know that marijuana users are *definitely* more likely to use other drugs later – for example, marijuana users are 85x more likely than non-marijuana users to use cocaine. but that could be either because marijuana affects them in some way (implying that legalizing marijuana would increase other drug use), because they have factors like genetics or stressful life situation that makes them more likely to use all drugs (implying that legalizing marijuana would not affect other drug use), or because using illegal marijuana without ill effect connects them to the illegal drug market and convinces them illegal drugs are okay (implying that legalizing marijuana

would decrease other drug use). RAND comes very close to investigating this properly by saying that when the Dutch pseudo-legalized marijuana, use of harder drugs stayed stable or went down, but all their study actually shows is that the ratio of marijuana users : hard drug users went down. This is to be expected when you make marijuana much easier to get, but it's still consistent with the absolute number of hard drug users going way up. The best that can be said is that there is no direct causal evidence for the gateway theory and some good alternative explanations for the effect. Let us accept their word for it and never speak of this matter again.

Conclusion for this section: Marijuana does not have a detectable effect on mortality and there is surprisingly scarce evidence of tobacco-like side effects. It probably does decrease IQ if used early and often, possibly by as many as 8 IQ points. It may increase risk of psychosis by as much as 40%, but it's not clear who is at risk or whether the risk is even real. The gateway drug hypothesis is too complicated to evaluate effectively but there is no clear casual evidence in its support.

III. What Are The Costs Of The Drug War?

There are not really that many people in jail for using marijuana.

I learned this from Who's Really In Prison For Marijuana?, a publication of the National Office Of Drug Control Policy, which was clearly written by someone with the same ability to take personal offense at bad statistics that inspires my posts about Facebook. The whole thing seethes with indignation and makes me want to hug the drug czar and tell him everything will be okay.

Only 1.6% of state prisoners are serving time for marijuana, only 0.7% are serving for marijuana possession, and only 0.3% are first time offenders. Some of those are "possession" in the sense of "possessing a warehouse full of marijuana bales", and others are people who committed much more dangerous crimes but were nailed for marijuana, in the same sense that Al Capone was nailed for tax evasion. The percent of normal law-abiding people who just had a gram or two of marijuana and were thrown in jail is a rounding error, and the stories of such you read in the news are extremely dishonest (read the document for examples).

Federal numbers are even lower; in the entire federal prison system, they could only find 63 people imprisoned with marijuana possession as the sole crime, and those people were possessing a median of one hundred fifteen *pounds* of marijuana (enough to make over 100,000 joints).

In total, federal + state prison and counting all the kingpins, dealers, manufacturers, et cetera, there are probably about 16,000 people in prison solely for marijuana-related offenses, serving average actual sentence lengths of three year. But it's anybody's guess whether those people would be free today if marijuana were legal, or whether their drug cartels would just switch to something else.

Looking at the other side's statistics, I don't see much difference. NORML claims that there are 40,000 people in prison for marijuana use, but they admit that half of those people were arrested for using harder drugs and marijuana was a tack-on charge, so they seem to agree with the Feds about around 20,000 pure marijuana prisoners. SAM agrees that only 0.5% of the prison population is in there for marijuana possession alone. I see no reason to doubt any of these numbers.

A much more serious problem is marijuana-related arrests, of which there are 700,000 a year. 90% of them are for simple possession, and the vast majority do not end in prison terms; they do however result in criminal records, community service, a couple days of jail time until a judge is available to hear the case, heavy fines, high cost of legal representation, and moderate costs to the state for funding the whole thing. Fines can be up to $1500, and legal representation can cost up to $5000 (though I am suspicious of this paper and think it may be exaggerating for effect). These costs are often borne by poor people who will have to give up all their savings for years to pay them back.

Costs paid by the government, which cover everything from police officers to trials to prison time, are estimated at about $2 billion by multiple sources. This is only 3% of the total law enforcement budget, so legalizing marijuana wouldn't create some kind of sudden revolution in policing, but as the saying goes, a billion here, a billion there, and eventually it adds up to real money. And a Harvard economist claims that the total monetary benefits from legalization, including potential tax revenues, could reach $14 billion.

Some people worry that legalizing marijuana would cause an increase in car accidents by "stoned drivers", who, like drunk drivers, have impaired reflexes and poor judgment, and indeed there is a small but real problem of marijuana-induced car accidents. But Chaloukpa and Laixuthai (1994) crunch the numbers and find that decreased price/increased availability of marijuana is actually associated with *decreased* car accidents, probably because marijuana is substituting for alcohol in the "have impairing substances and then go driving" population. This finding – that marijuana and alcohol substitute for each other – has been spotted again and again. Anderson & Rees

(2013) find that states that legalize medical marijuana see a 5% drop in beer sales. There are however a few dissenting opinions: Cameron & Williams (2001), in complex econometric simulations that may or may not resemble the real world in any respect, find that increasing the price of alcohol increases marijuana use, but increasing the price of marijuana does not affect alcohol use, and the same researcher finds that banning alcohol on a college campus also decreases marijuana use. Also, possibly marijuana use increases smoking? This whole area is confusing, but I am most sympathetic to to the Andersen and Rees statistics which say that medical marijuana states are associated with 13% fewer traffic fatalities.

Overall conclusion for this section: full legalization of marijuana would free about 20,000 people from jail (although most of them would not be exactly fine upstanding citizens), prevent 700,000 arrests not resulting in jail time per year, save between 2 and 14 billion dollars, and possibly reduce traffic fatalities a few percent (or, for all we know, increase them).

IV. An Irresponsible Utilitarian Analysis

Decriminalization and legalization of medical marijuana seem, if we are to trust the statistics in (I) saying they do not increase use among youth, like almost unalloyed good things. Although there are some nagging hints of doubt, they are not especially quantifiable and therefore not amenable to analysis. Without a very strong predisposition to try as hard as possible to fit the evidence into a pessimistic picture, I don't think there's a great argument against either of these two propositions. Let's concentrate on legalization, which would mean something like "People can grow and sell as much marijuana as they want and it's totally legal for people over 21, with the same level of penalties as today for people under 21".

Section (I) concludes that legalization could lead to an increase in adult marijuana use up to 50%. There's not a lot of evidence on what it could do to teen marijuana use, but since it seems teen marijuana use is less responsive to legal changes, I made up a number and said 20%. Lest you think I am being unfair, note that this is well below the percent increase predicted by the survey that asked 18 year olds if they would start using marijuana if it were legal.

Right now about 1.5 million teenagers use marijuana "heavily". Most of the detrimental effects of marijuana seem concentrated in teens and people in their early twenties; I'm going to artificially round that up to 2 million to catch the early 20 year olds. If this 2 million number increased 20%, 400,000 extra teens would start heavily using marijuana.

Those 400,000 teens would lose 8 IQ points each. IQ increases your yearly earnings by about $500 per point, so these people would lose about $4,000 a year. Making very strong assumptions about salary being a measure of value to society, society would lose about $1.6 billion a year directly, plus various intangibles from potential artists and scientists losing the ability to create masterpieces and inventions, plus various *really* intangibles like a slightly dumber electorate.

We need to use a different number to calculate psychosis risk, since the studies were done on "people who had used marijuana at least once". The appropriate number turns out to be 8 million teenagers; of those, 1%, or 80,000, would naturally develop schizophrenia. If the 1.4 relative risk number is correct, marijuana use will increase that to 112,000, for a total increase of 32,000 people. Schizophrenia pretty much always presents in the 15 – 25 age window, so we'll say we get 3,200 extra cases per year.

There were 35000 road traffic accident fatalities in the US last year. If greater availability of marijuana decreases those fatalities by 13% (note that I am using the number from medical marijuana legalization and not for marijuana legalization per se, solely because it is a number I actually have), that will cause 4500 fewer road traffic deaths per year. There may be additional positive effects of alcohol substitution from, for example, less liver disease. But there may also be additional negative effects from increasing use of tobacco, so let's just pretend those cancel out.

So here is my guess at the yearly results of marijuana legalization:

– 20,000 fewer prisoners (but they might switch to other criminal enterprises)
– 700,000 fewer arrests
– $2 billion less in law enforcement costs
– Some amount of positive gain (let's say $5 billion) in taxes
– 4500 fewer road traffic deaths (if you believe the preliminary alcohol substitution numbers)

– 400,000 people with lower IQ
– $2 billion in social costs from above dumber people
– 3,200 more cases of schizophrenia a year

We'll proceed to calculate the nonmonetary burden of each of these in QALYs, then add the monetary burden in dollars, then convert.

The searchable public database of utility weights for all diseases (God I love the 21st century) tells me that schizophrenia has a QALY weight of 0.73. It generally starts around 20 and lasts a lifetime, so each case of schizophrenia costs us 0.27 * 50 or 13.5 QALYs. Therefore, the total burden of the 3,200 added schizophrenia cases is 43 kiloQALYs.

There's no good way to calculate the QALY weight of having 4-8 fewer IQ points, and unfortunately this is going to end up being among the most important numbers in our results. If we say the lifetime cost of this problem is 3 QALYs, and divide the number by eight to represent eight years worth of teenagers in our sample population, we end up with 400,000/8 * 3 = 150 kiloQALYs.

My own survey tells me that being in prison has a QALY weight around 0.5. Marijuana sentences generally last an average of three years, which suggests that 1/3 of these marijuana prisoners are arrested every year, so the total burden of the ~6000ish marijuana imprisonments each year is 3 * ~6000 * 0.5 = 10 kiloQALYs.

Assume the average road traffic death occurs at age 30, costing 40 years of potential future life. The total cost of 4500 road traffic deaths is 40 * 4500 = 180 kiloQALYs.

The arrests are going to require even more fudging than normal. Average jail time for a marijuana arrest (when awaiting trial) is "one to five days" – let's round that off to two and then use our prison number to say that the jail from each arrest is 2/365 * 0.5 = three-thousandths of a QALY. I am going to arbitrarily round this up to one one-hundredth of a QALY to account for emotional trauma and the burden of fines, then even more arbitrarily round this up to a tenth of a QALY to account for possibility of getting a criminal record. This sets the burden of 700,000 arrests at 70 kiloQALYs.

Now our accounting is:

Costs from legalization compared to current system: 200 kQALYs and $2 billion
Benefits from legalization compared to current system: 260 kQALYs and $7 billion

Although it's not going to be necessary, we can interconvert QALYs and dollars at the going health-care rate of about $100,000/QALY ($100 million/kQALY):

Costs from legalization compared to current system: 220 kQALYs
Benefits from legalization compared to current system: 330 kQALYs

And get:

Net benefits from legalization: +110 kQALYs

Except that this is extremely speculative and irresponsible. By far the largest component of the benefits of legalization turned out to be the effect on road traffic accidents, which is based on only two studies and which may on further research turn out to be a cost. And by far the largest component of the costs of legalization turned out to be the effect on IQ, and we had to totally-wild-guess the QALY cost of an IQ point loss. The wiggle room in my ignorance and assumptions is more than large enough to cover the small gap between the two policies in the results.

So my actual conclusion is:

There is not a sufficiently obvious order-of-magnitude difference between the costs and benefits of marijuana legalization for a evidence-based utilitarian analysis of costs and benefits to inform the debate. You may return to your regularly scheduled wild speculation and shrill accusations.

But I wouldn't say this exercise is useless. For example, it suggests that whether marijuana legalization is positive or negative on net depends almost entirely on small changes in the road traffic accident rate. This is something I've never heard anyone else mention, but which in retrospect should be

obvious; the few debatable health effects and the couple of people given short jail sentences absolutely can't compare to the potential for thousands more (or fewer) traffic accidents which leave people permanently dead.

So my actual actual conclusion is:

We should probably stop caring about health effects of marijuana and about imprisonment for marijuana-related offenses, and concentrate all of our research and political energy on how marijuana affects driving.

This cements my previous intuitions on irresponsible use of statistics – it's unlikely to unilaterally solve the problem, but it can be very good at pointing out where you're being irrational and suggesting new ways of looking at a question.

EDIT: People in the comments have pointed out several important factors left out, including:
– Some people enjoy smoking marijuana
– The opening of a permanent criminal record may mean arrests are worse than I estimate. I can't find good statistics on how often this happens, but do note that decriminalization prevents a record from being opened.
– Loss of 8 IQ points may have wider social effects than I estimate, since IQ affects for example crime rate.
– Legalizing marijuana might remove a source of funding for organized crime

Are You a Solar Deity?

Max Muller was one of the greatest religious scholars of the 19th century. Born in Germany, he became fascinated with Eastern religion, and moved to England to be closer to the center of Indian scholarship in Europe. There he mastered English and Sanskrit alike to come out with the first English translation of the Rig Veda, the holiest book of Hinduism.

One of Muller's most controversial projects was his attempt to interpret all pagan mythologies as linked to one another, deriving from a common ur-mythology and ultimately from the celestial cycle. His tools were exhaustive knowledge of the myths of all European cultures combined with a belief in the interpretive power of linguistics.

What the significance of Orpheus' descent into the underworld to reclaim his wife's soul? The sun sets beneath the Earth each evening, and returns with renewed brightness. Why does Apollo love Daphne? Daphne is cognate with Sanskrit Dahana, the maiden of the dawn. The death of Hercules? It occurs after he's completed twelve labors (cf. twelve signs of zodiac) when he's travelling west (like the sun), he is killed by Deianeira (compare Sanskrit dasya-nari, a demon of darkness) and his body is cremated (fire = the sun). His followers extended the method to Jesus - who was clearly based on a lunar deity, since he spent three days dead and then returned to life, just as the new moon goes dark for three days and then reappears.

Muller's work was massively influential during his time, and many 19th century mythographers tried to critique his paradigm and poke holes in it. Some accused him of trying to

destroy the mystery of religion, and others accused him of shoddy scholarship.

R.F. Littledale, an Anglican apologist, took a completely different route. He claimed that there was, in fact, no such person as Professor Max Muller, holder of the Taylorian Chair in Modern European Languages. All these stories about "Max Muller" were nothing but a thinly disguised solar myth.

Littledale begins his argument by noting Muller's heritage. He was supposedly born in Germany, only to travel to England when he came of age. This looks suspiciously like the classic Journey of the Sun, which is born in the east but travels to the west. Muller's origin in Germany is a clear reference to Germanus Apollo, one of the old appelations of the Greek sun god.

His Christian name must be related to Latin "maximus" or Sanskrit "maha", meaning great, a suitable description of the King of Gods, and his surname is cognate with Mjolnir, the mighty hammer of the sky god Thor. His claim to fame is bringing the ancient wisdom of the East to the people of the West - that is, illuminating them with eastern light.

Muller teaches at Oxford for the same reason that Genesis describes the sky as "the waters above" and the Egyptians gave Ra a solar barge: ancient people interpreted the sky as a river, and the sun as crossing that river upon his chariot (perhaps an **ox**-drawn chariot, **ford**ing the river?). His chair at Oxford is the throne of the sky, his status as Taylorian Professor because "he cuts away with his glittering shears the ragged edges of cloud; he allows the…cuttings from his workshop, to descend in fertilizing showers upon the earth."

I could go on; instead I recommend you read the original essay. The take-home lesson is that any technique powerful

enough to prove that Hercules is a solar myth is also powerful enough to prove that *anyone* is a solar myth. Muller lacked the strength of a rationalist: the ability to be more confused by fiction than by reality. This makes the Hercules theory useless, but that is not immediately apparent on a first or even a second reading of Muller's work. When reading Muller's work, the primary impression one gets is "Wow, this man has gathered a *lot* of supporting evidence."

This is a problem encountered in many fields of scholarship, especially "comparative" anything. In comparative linguistics, for example, it's usually possible to make a case that two languages are related good enough to convince a layman, no matter which two languages or how distant they may be. In comparative religion, we get cases like this blog's recent discussion over the possible derivation of Esther and Mordechai defeating Haman from Ishtar and Marduk defeating Humbaba. The less said about comparative literature, the better, although I can't help but quote humor writer Dave Barry:

> Suppose you are studying Moby-Dick. Anybody with any common sense would say that Moby-Dick is a big white whale, since the characters in the book refer to it as a big white whale roughly eleven thousand times. So in *your* paper, *you* say Moby-Dick is actually the Republic of Ireland. Your professor, who is sick to death of reading papers and never liked Moby-Dick anyway, will think you are enormously creative. If you can regularly come up with lunatic interpretations of simple stories, you should major in English.

The worst (but most fun to read!) are in pseudoscience, where plausible sounding comparisons can prove almost anything.

Did you know the Mayans believed in a lost homeland called Atzlan, the Indonesians believed in a lost island called Atala, and the Greeks believed in a lost continent called Atlantis? Likewise, did you know that Nostradamus predicted a great battle involving Germany and "Hister", which sounds almost like "Hitler"?

Yet it would be a mistake to reject all such comparisons. In fact, I have thus far been enormously unfair to Professor Muller, whose work established several correspondences still viewed as valid today. Virtually all modern mythologists accept that the Hindu Varuna is the Greek Uranus, and that the Greek sky god Zeus equals the Hindu sky god Dyaus Pita and the Roman Jupiter (compare to Latin *deus pater*, meaning God the Father). Likewise, comparative linguists are quite certain that all modern European languages and Sanskrit derive from a common Indo-European root, and in my opinion even the Nostratic project - an ambitious attempt to link Semitic, Indo-European, Uralic. and a bunch of other languages - is at least worth consideration.

We need a test to distinguish between true and false correspondences. But the standard method, making and testing predictions, is useless here. A good mythologist already knows the stories of Varuna and Uranus. The chances of discovering a new fact that either confirms or overturns the Varuna-Uranus correspondence is not even worth considering.

Mark Rosenfelder has an excellent article on chance resemblances between languages which offers a semi-formal model for spotting dubious comparisons. But such precision may not be possible when comparing two deities.

I have what might be a general strategy for approaching this sort of problem, which I will present tomorrow. But how

would you go about it?

The “Spot the Fakes” Test

Followup to: Are You a Solar Deity?

James McAuley and Harold Stewart were mid-20th century Australian poets, and they were not happy. After having society ignore their poetry in favor of “experimental” styles they considered fashionable nonsense, they wanted to show everyone what they already knew: the Australian literary world was full of empty poseurs.

They began by selecting random phrases from random books. Then they linked them together into something sort of like poetry. Then they invented the most fashionable possible story: Ern Malley, a loner working a thankless job as an insurance salesman, writing sad poetry in his spare time and hiding it away until his death at an early age. Posing as Malley’s sister, who had recently discovered the hidden collection, they sent the works to Angry Penguins, one of Australia’s top experimental poetry magazines.

You wouldn’t be reading this if the magazine hadn’t rushed a special issue to print in honor of “a poet in the same class as W.H. Auden or Dylan Thomas”.

The hoax was later revealed[1], everyone involved ended up with egg on their faces, and modernism in Australia received a serious blow. But as I am reminded every time I look through a modern poetry anthology, one Ern Malley every fifty years just isn’t enough. I daydream about an alternate dimension where people are genuinely interested in keeping literary criticism honest. In this universe, any would-be literary critic would have to distinguish between ten poems generally recognized as brilliant that he’d never seen before, and ten

pieces of nonsense invented on the spot by drunk college students, in order to keep his critic's license.

Can we refine this test? And could it help Max Muller with his solar deity problem?

In the Malley hoax, McAuley and Steward suspected that a certain school of modernist poetry was without value. Because its supporters were too biased to admit this directly, they submitted a control poem they knew was without value, and found the modernists couldn't tell the difference. This suggests a powerful technique for determining when something otherwise untestable might be, as Neal Stephenson calls it, bulshytte.

Perhaps Max Muller thinks Hercules is a solar deity. He will write up a argument for this proposition, and submit it for consideration before all the great mythologists of the world. Even if these mythologists want to be unbiased, they will have a difficult time of it: Muller has a prestigious reputation, and they may not have any set conception of what does and doesn't qualify as a solar deity.

What if, instead of submitting one argument, Muller submitted ten? One sincere argument for why Hercules is a solar deity, and other bogus arguments for why Perseus, Bellerophon, Theseus, et cetera are solar myths (which he has nevertheless constructs to the best of his ability). Then he instructs the mythologists "Please independently determine which of these arguments is true, and which ones I have just come up with by writing 'X is a solar deity' as my bottom line and then inventing fake justifications for the fact?" If every mythologist finds the Hercules argument most convincing, then that doesn't prove anything about Hercules but it at least shows Muller has a strong case. On the other hand, if they're all

convinced by different arguments, or find none of the arguments convincing, or worst of all they all settle on Bellerophon, then Dr. Muller knows his beliefs about Hercules are quite probably wishful thinking.

This method hinges on Dr. Muller's personal honesty: a dishonest man could simply do a bad job arguing for Theseus and Bellerophon. What if we thought Dr. Muller was dishonest? We might find another mythologist whom independent observers rate as equally persuasive as Dr. Muller, and ask her to come up with the bogus arguments.

The rationalists I know sometimes take a dim view of the humanities as academic disciplines. Part of the problem is the seeming untestability of their conclusions through good, blinded experimental methods. I don't think most humanities professors are really looking all that hard for such methods. But for those who are, I consider this technique a little better than nothing[2].

Footnotes

1: The Sokal Affair is another related hoax. Wikipedia's Sokal Hoax page has some other excellent examples of this sort of test.

2: One more example where this method could prove useful. I remember debating a very smart Christian on the subject of Biblical atrocities. You know, stuff about death by stoning for minor crimes, or God ordering the Israelites to murder women and enslave children - that sort of thing. My friend, who was quite smart, was always able to come up with a superficially plausible excuse, and it was getting on my nerves. But having just read Your Strength as a Rationalist, I knew that being able to explain anything wasn't always a virtue. I proposed the following experiment: I'd give my friend ten atrocities

commanded by random Bronze Age kings generally agreed by historical consensus to be jerks, and ten commanded by God in the Bible. His job would be to determine which ten, for whatever reason, really weren't all that bad. If he identified the ten Bible passages, that would be strong evidence that Biblical commandments only seemed atrocious when misunderstood. But if he couldn't tell the difference between God and Ashurbanipal, that would prove God wasn't really that great. To my disgust, my friend knew his Bible so well that I couldn't find any atrocities he wasn't already familiar with. So much for that technique. I offer it to anyone who debates theists with less comprehensive knowledge of Scripture.

Epistemic Learned Helplessness

[**Epistemic Status** | *Probably I'm just coming at the bog-standard idea of compartmentalization from a different angle here. I don't know if anyone else has noted how compartmentalization is a good thing before, but I bet they have.*]

A friend in business recently complained about his hiring pool, saying that he couldn't find people with the basic skill of *believing arguments*. That is, if you have a valid argument for something, then you should accept the conclusion. Even if the conclusion is unpopular, or inconvenient, or you don't like it. He told me a good portion of the point of CfAR was to either find or create people who would believe something *after it had been proven to them*.

And I nodded my head, because it sounded reasonable enough, and it wasn't until a few hours later that I thought about it again and went "Wait, no, that would be the worst idea ever."

I don't think I'm overselling myself too much to expect that I could argue circles around the average high school dropout. Like I mean that on almost any topic, given almost any position, I could totally demolish her and make her look like an idiot. Reduce her to some form of "Look, everything you say fits together and I can't explain why you're wrong, I just *know you are*!" Or, more plausibly, "Shut up I don't want to talk about this!"

And there are people who can argue circles around me. Not on any topic, maybe, but on topics where they are experts and have spent their whole lives honing their arguments. When I was young I used to read pseudohistory books; Immanuel Velikovsky's Ages in Chaos is a good example of the best this genre has to offer. I read it and it seemed so obviously correct,

so *perfect*, that I could barely bring myself to bother to search out rebuttals.

And then I read the rebuttals, and they were so obviously correct, so *devastating*, that I couldn't believe I had ever been so dumb as to believe Velikovsky.

And then I read the rebuttals to the rebuttals, and they were so obviously correct that I felt silly for ever doubting.

And so on for several more iterations, until the labyrinth of doubt seemed inescapable. What finally broke me out wasn't so much the lucidity of the consensus view so much as starting to sample different crackpots. Some were almost as bright and rhetorically gifted as Velikovsky, all presented insurmountable evidence for their theories, and all had mutually exclusive ideas. After all, Noah's Flood couldn't have been a cultural memory *both* of the fall of Atlantis *and* of a change in the Earth's orbit, let alone of a lost Ice Age civilization or of megatsunamis from a meteor strike. So given that at least some of those arguments are wrong and all seemed practically proven, I am obviously just gullible in the field of ancient history. Given a total lack of independent intellectual steering power and no desire to spend thirty years building an independent knowledge base of Near Eastern history, I choose to just accept the ideas of the prestigious people with professorships in Archaeology rather than the universally reviled crackpots who write books about Venus being a comet.

I guess you could consider this a form of *epistemic learned helplessness*, where I know any attempt to evaluate the arguments are just going to be a bad idea so I don't even try. If you have a good argument that the Early Bronze Age worked completely differently from the way mainstream historians believe, I *just don't want to hear about it*. If you insist on

telling me anyway, I will nod, say that your argument makes complete sense, and then totally refuse to change my mind or admit even the slightest possibility that you might be right.

(This is the correct Bayesian action, by the way. If I know that a false argument sounds just as convincing as a true argument, argument convincingness provides no evidence either way, and I should ignore it and stick with my prior.)

I consider myself lucky in that my epistemic learned helplessness is circumscribed; there are still cases where I will trust the evidence of my own reason. In fact, I trust it in most cases other than very carefully constructed arguments known for their deceptiveness in fields I know little about. But I think the average high school dropout both doesn't and *shouldn't*. Anyone anywhere - politicians, scammy businessmen, smooth-talking romantic partners - would be able to argue her into anything. And so she takes the obvious and correct defensive manuever - she will never let anyone convince her of any belief that sounds "weird" (note that, if you grow up in the right circles, beliefs along the lines of astrology *not* working sound "weird".)

This is starting to sound a lot like ideas I've already heard centering around compartmentalization and taking ideas seriously. The only difference between their presentation and mine is that I'm saying that for 99% of people, 99% of the time, *this is a terrible idea*. Or, at the very least, this should be the *last* skill you learn, after you've learned every other skill that allows you to know which ideas are or are not correct.

The people I know who are best at taking ideas seriously are those who are smartest and most rational. I think people are working off a model where these co-occur because you need to be very clever to fight your natural and detrimental tendency

not to take ideas seriously. I think it's at least possible they co-occur because you have to be *really smart* in order for taking ideas seriously to be even not-immediately-disastrous. You have to be really smart not to have been talked into enough terrible arguments to develop epistemic learned helplessness.

Even the smartest people I know have a commendable tendency not to take certain ideas seriously. Bostrom's simulation argument, the anthropic doomsday argument, Pascal's Mugging - I've never heard anyone give a coherent argument against any of these, but I've also never met anyone who fully accepts them and lives life according to their implications.

A friend tells me of a guy who once accepted fundamentalist religion because of Pascal's Wager. I will provisionally admit that this person takes ideas seriously. Everyone else loses.

Which isn't to say that some people don't do better than others. Terrorists seem pretty good in this respect. People used to talk about how terrorists must be very poor and uneducated to fall for militant Islam, and then someone did a study and found that they were disproportionately well-off, college educated people (many were engineers). I've heard a few good arguments in this direction before, things like how engineering trains you to have a very black-and-white right-or-wrong view of the world based on a few simple formulae, and this meshes with fundamentalism better than it meshes with subtle liberal religious messages.

But to these I would add that a sufficiently smart engineer has never been burned by arguments above his skill level before, has never had any reason to develop epistemic learned helplessness. If Osama comes up to him with a really good argument for terrorism, he thinks "Oh, there's a good

argument for terrorism. I guess I should become a terrorist," as opposed to "Arguments? You can prove *anything* with arguments. I'll just stay right here and not do something that will get me ostracized and probably killed."

Responsible doctors are at the other end of the spectrum from terrorists in this regard. I once heard someone rail against how doctors totally ignored all the latest and most exciting medical studies. The same person, practically in the same breath, then railed against how 50% to 90% of medical studies are wrong. *These two observations are not unrelated.* Not only are there so many terrible studies, but pseudomedicine (not the stupid homeopathy type, but the type that links everything to some obscure chemical on an out-of-the-way metabolic pathway) has, for me, proven much like pseudohistory in that unless I am an expert in *that particular field of medicine* (biochemistry has a disproportionate share of these people and is also an area where I'm weak) it's hard not to take them seriously, even when they're super-wrong.

I have developed a healthy dose of epistemic learned helplessness, and the medical establishment offers a shiny tempting solution - first, a total unwillingness to trust anything, no matter how plausible it sounds, until it's gone through an endless cycle of studies and meta-analyses, and second, a bunch of Institutes and Collaborations dedicated to filtering through all these studies and analyses and telling you what lessons you should draw from them. Part of the reason *Good Calories, Bad Calories* was so terrifying is that it made a strong case that this establishment can be very very wrong, and I don't have good standards by which to decide whether to dismiss it as another Velikovsky, or whether to just accept that the establishment is totally untrustworthy and, as doctors sometimes put it, AMYOYO. And if the latter, how much

establishment do I have to jettison and how much can be saved? Do I have to actually go through all those papers purporting to prove homeopathy with an open mind?

I am glad that some people never develop epistemic learned helplessness, or develop only a limited amount of it, or only in certain domains. It seems to me that although these people are more likely to become terrorists or Velikovskians or homeopaths, they're also the only people who can figure out if something basic and unquestionable is wrong, and make this possibility well-known enough that normal people start becoming willing to consider it.

But I'm also glad epistemic learned helplessness exists. It seems like a pretty useful social safety valve most of the time.

III. Science and Doubt

Google Correlate Does Not Imply Google Causation

I need something sexy, something to lure new readers to this new blog and get them excited. So let's talk about statistical correlations. No, wait, *failed* statistical correlations!

Google Correlate is a nifty new Google product that takes data sets and finds search terms that correlate with them. For example, if you set it to "correlate over time" and enter a data set of average US temperature, it might return the search term "skiing", because people are most likely to ski when it's cold and so searches for skiing will be correlated with temperature. You can also just enter in Google search terms and see what *other* search terms they're correlated with.

The results seem to fall into two categories: obvious and nonsensical.

Ones with clear time patterns are obvious. If you enter in skiing, you'll get "how to ski", "buy skis", "snowboarding", "ski resorts", and the like. If you enter in a news trend that was only popular at one point, you'll get both related terms and other news trends only popular at that one point – for example, "school shooting" brings up "jan berenstain", not because the Berenstain Bears books secretly cause school shootings (…one hopes) but because she died the same week as a relatively big one and so people were searching them around the same time.

Things that don't have obvious time patterns seem to bring up results that are both nonsensical and very-very convincing-looking. The worst are diseases.

This is Google Correlate's result for *heart attack*. It matches it to "pink lace dress" with a correlation of .88 (for comparison,

a study comparing cigarette use vs. lung cancer rates across different social groups found a correlation of .71).

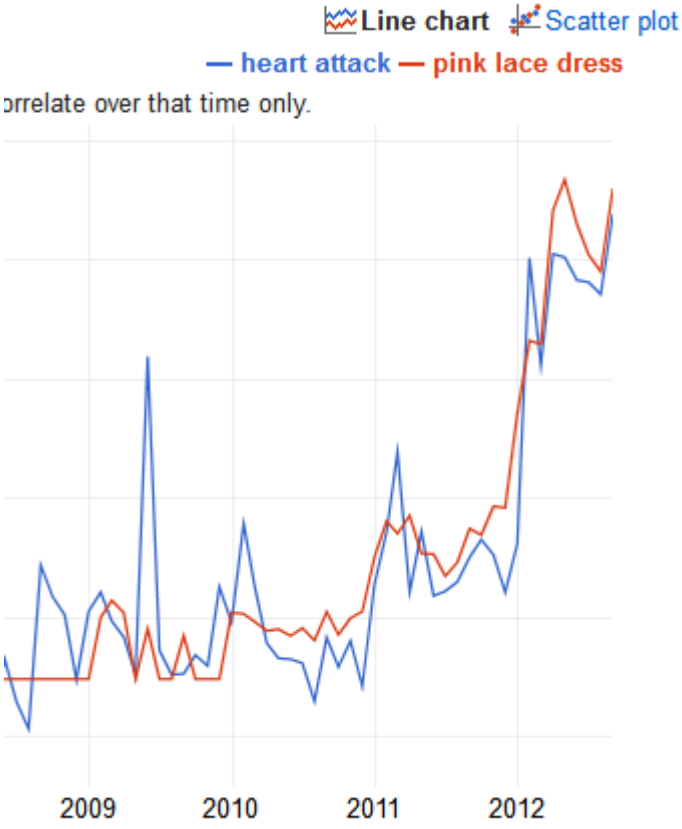


Figure 1: Correlation between interest in heart attacks and in pink lace dresses, by time.

As far as I can tell, this is just an artifact of Google having lots and lots of search terms and you would expect some of them to be heavily correlated by mere coincidence.

Google also has a correlate-by-state feature. This one has even weirder results for heart attack, like "can you get a" and "is it a" (note that these are the entire search terms). I understand that "is it a heart attack" is a reasonable question, but I don't understand who would just enter that phrase into Google and hope it would figure it out. I'm kind of imagining someone having a heart attack going on Google, typing as far as "is it a…" and then falling over dead, but I assume the real

explanation is more prosaic, like someone expecting autocomplete to work but being disappointed.

Google's state-by-state feature seemed potentially really exciting to me. I wrote a while back on the effect of parasite load, and I had the dataset lying around with different states ranked on different metrics. I entered the data for parasite load and got the following search terms: "Toy Johnson", "Bernie Mac", "booty models", "Harvey suits", "Beyonce clothing line".

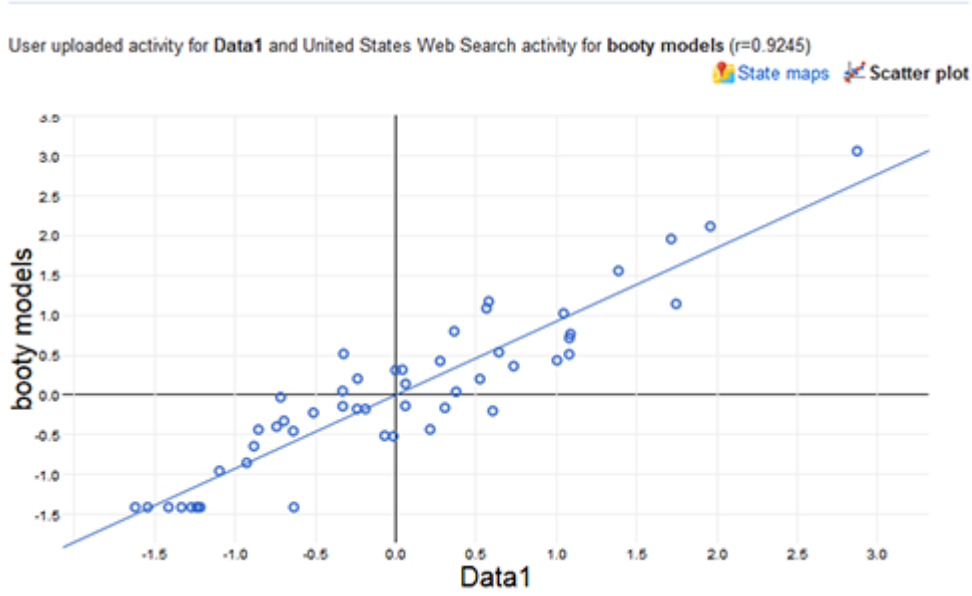


Figure 2: Correlation between parasite prevalence and interest in booty models, by state.

I didn't actually know what most of these were (I kinda thought Bernie Mac was a real estate conglomerate, which turns out to be false) but upon closer investigation they are all black people or Stuff Black People Like. So I think what's happening here is that the high-parasite load states are all in the South and relatively poor with low access to health care, which also selects for black people. This obviously has significant implications for the study's attempt to determine that high parasite load causes certain social trends.

My next thought was “if I multiply this data set by negative one, I will have an objective pipeline to figuring out Stuff White People Like. That sounds interesting.” So I tried it, and my results were: “black albino”, “shake that eminem”, “tony hawk pro skater”, and “green day time of your life”. I was sort of hoping that “Black Albino” was the name of a band or something (it would actually be a pretty good one) but no, it turns out white people are just fascinated with the idea of black albinos. White people are kind of weird.

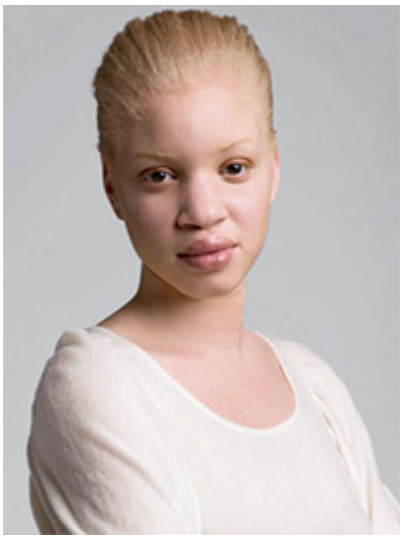


Figure 3: A black albino. Happy now, white people?

But let’s keep going through the state-by-state data set. My next Big Social Statistic was “importance of family ties, by state”. States with higher family ties were more likely to search for: “how to swim”, “composition book”, “noni juice”, “muscle men”, “girl kiss”, “Toyota Tacoma 2008”.

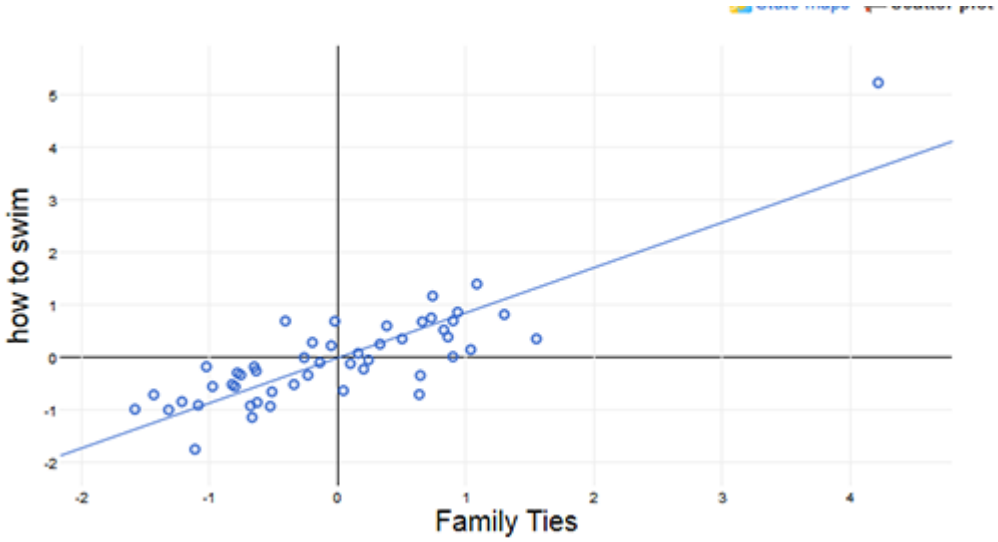


Figure 4: Correlation between strength of family ties and interest in swimming, by state.

A lot of these seem related to physical fitness, or ruggedness (the Tacoma seems to be a very sporty, rugged car), or masculinity. I'm not really sure what to make of this.

The last Social Science Statistic in the dataset was Religiosity, which correlated with the following search terms: "Christmas themes", "rotary cutter", "Honda rebel 250". Christmas themes seems sort of plausible. I dunno about the rest.

So as far as I can tell Google Correlate is not very interesting. It doesn't reveal any deep connections between concepts, or even guess what concept my dataset came from to begin with. For something potentially so powerful this is disappointing.

I can think of two possible uses for it. The first is as a sanity check to make sure your data aren't completely confounded. If you think you're measuring average number of roof tiles per house or something, and your data's Google Correlate results come back with Toy Johnson and Beyonce clothing, you're probably just measuring race and for some reason different races have different numbers of roof tiles on their houses. Which means if you think you've found a correlation between

roof tiles and something fascinating like voting record, you're probably just being confounded by race. This is a real problem in a lot of studies.

The second is as a cheap hack for *creating* datasets. I entered "Jesus" in and got a state by state list of who searched for Jesus. It looked a lot like my state-by-state map of religiousity. The correlates were all things like "Apostle", "Paul", "preaching", and for some reason "Abednego", who is a very minor Biblical character who has no business being in the top ten correlates of Jesus at all. If you wanted to make a cheap map of state-religiosity in order to correlate to parasite load or whatever, Google Trends seems like a plausible method.

On the other hand, I tried to see if I could recreate their state map of parasite load. I asked it to correlate "metronidazole", a medication commonly used in the treatment of parasitic diseases, on the grounds that people with parasites would be prescribed metronidazole and then look it up to see if it was safe. The result looked only a little like my map of state-by-state parasite data, and the number one correlated search term (r = .89) was "Is Lil' Wayne gay?"

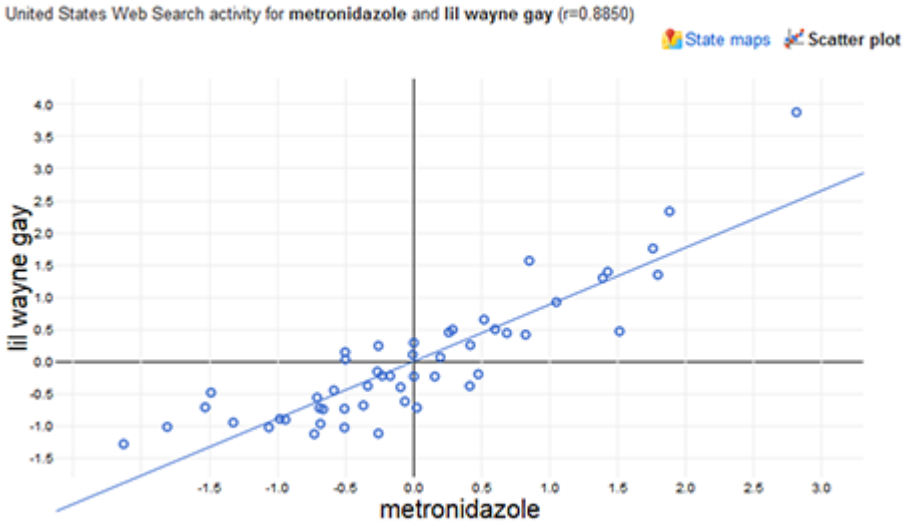


Figure 5: Correlation between curiosity over Lil' Wayne's sexual orientation and interest in the anti-parasitic medication metronidazole. Whatever my case was, I hereby rest it.

So if nothing else, this exercise has proven my suspicion that the sort of people who worry about whether Lil' Wayne is gay are, in fact, crawling with parasites.

Stop Confounding Yourself! Stop Confounding Yourself!

As a perk of my job, I get a free subscription to the *American Journal of Psychiatry*. I am still not used to this. No enraging struggles with paywalls. No "one year embargo on full text". I just come home and find all of the latest and most interesting journal articles have been *shipped directly to my house*. Modern technology is truly amazing.

Its latest is Takizawa et al's **Adult Health Outcomes of Childhood Bullying Victimization: Evidence From A Five-Decade Longitudinal British Birth Cohort**. It has since been picked up by Fox, the Washington Post, and even Xinhua. I think that's enough to qualify for "made world headlines".

The study took some British kids in 1958, sorted them by how much they got bullied, and checked how they did forty years later. In fact, the frequently bullied kids had nearly twice as much psychiatric disease, were twice as likely to attempt suicide, were twice as likely to drop out of high school, and even had double the unemployment rate. Worse physical health, worse cognitive function, less likely to get married, et cetera, et cetera.

Those must be *some* bullies.

But correlation is not causation. There's an alternative possibility. Maybe bullies only pick on unpopular disadvantaged kids. And maybe these kinds of things are stable, so that unpopular disadvantaged kids are more likely to grow up to be unpopular disadvantaged adults. The sort of adults who are more likely to have psychiatric disease, drop

out of school, be unemployed, et cetera. *That* sure sounds plausible.

So the researchers "controlled for confounders". They used a scale called the Bristol Social Adjustment Guide to figure out how socially well-adjusted the kids were, then aded in their social class, their family's level of contact with child protective services, their IQ, their attractiveness, and even how much their parents loved them (really! check the study!)

They controlled for all these things and found that the bullying-outcomes link was still robust. They concluded that this meant their finding wasn't just that bullies were bullying kids with problems, it was that bullies were causing the damage themselves.

Do you believe that? It all comes down to one question.

Who is better able to look deep inside you and judge the mettle of your soul? A playground bully? Or the Bristol Social Adjustment Guide?

My money is on the bully. Bullies are like sharks: horrible pinnacles of evolution. Animals have been learning to navigate social dominance hierarchies through violence since pecking orders in chickens, on through wolf packs and chimpanzees, and up into humans – and we are very good at it. The bully is the purest manifestation of the primal instinct, which is why he crops up untaught and unbidden in near-identical form in schoolyards from Los Angeles to London to Lanzhou. And like sharks, a good bully should be able to smell blood in the water and know when an opportunity to attack presents itself.

Most of the findings of this study were in the "frequently bullied" population, and part of the criteria for "frequently" was bullying both at age 7 and age 11. Unless that's just one *really* persistent guy, that means the child has gotten

independently selected for targeting in two different environments. That could be bad luck but could also be the effect of high inter-bully reliability in what (persistent) qualities make a good victim.

So let's take another look at those confounders we supposedly controlled for. Where's height? You think short kids are bullied more often than tall kids? I do. Height is closely related to career success, to attractiveness to the opposite sex, increased happiness and self-esteem, and decreased psychological morbidity. This is something every bully knows intuitively, but which the Takizawa study didn't think of and therefore couldn't control for.

But it's giving them too much credit to be bringing in weird stuff like height-mental-health correlations. What about social skills? Yeah, sure, they did that Bristol Social Adjustment Guide. I'm looking at it right now, and it's asking the students' teachers to rate items like "hostility towards adults" and "depression". I don't believe that teachers filling numbers into hokey little boxes can capture an assessment of a kid's social skills as well as a bully trying to decide who can safely be picked on can.

So I will come out and say it: I do not trust the practice of "adjusting for confounders", at least not the way this study does it. You are adjusting for an imperfect measurement of the confounders you can think of. If you find that there is lingering correlation, then either you your hypothesis is true, or you didn't adjust for confounders well enough. Given extraordinary results, like being bullied at age seven making you 25% less likely to be married at age fifty, the "you didn't adjust for confounders well enough" option starts to look really good.

I think the proper way to do this study would have been to do an anti-bullying intervention at a couple schools, leave a couple similar schools as controls, and if the anti-bullying intervention successfully decreases bullying, compare outcomes for children at the two schools. I understand this probably would be logistically impossible, plus you'd have to wait another forty years. But given that you cannot do the study right, I am not sure that doing the study this way adds anything, except of course widely-read articles in every news source in the world.

I would also compare to Reming et al, which attempts much the same study and finds no association after adjusting for *their* confounders of choice (which, oddly, are much fewer than in the current study). They also find that parent reports about bullying (the method Takizawa et al used) are wildly unreliable, with an inter-rater agreement of just 0.11 with reports by teachers or the children themselves (the statistic goes from perfect agreement being 1.0 to zero information being 0.0). For a *completely false* measure of bullying to find such spectacular effects is *really suspicious*, and now we need to consider not only the differences between the types of kids who are and aren't bullied, but the differences between the types of parents who do and don't think their kids are being bullied.

Since I insisted on giving this post a silly title, I will now share with you the most interesting perspective on psychology and the "stop hitting yourself" phenomenon I have read all week. This is from Jonathan Haidt on Kohlberg's moral stages:

> During elementary school, most children move on to the two conventional stages, becoming adept at understanding and even manipulating rules and social

> conventions. This is the age of petty legalism that most of us who grew up with siblings remember well ("I'm not hitting you. I'm using your hand to hit you. Stop hitting yourself!"). Kids at this stage rarely question the legitimacy of authority, but learn to maneuver within and around the constraints that adults impose on them.

I always just thought that was a really dickish joke. I didn't realize it had a *deep philosophical underpinning*.

Effects of Vertical Acceleration on Wrongness

Whenever someone sneers “Evidence-based medicine? You wouldn’t demand a double-blind placebo-controlled clinical trial of PARACHUTES, would you?” I feel a strong urge to use them as the control group in my double-blind parachute experiment.

Of course, deep down inside I know that this would be morally wrong. Groups need to be determined by random assignment.

90% Of All Claims About The Problems With Medical Studies Are Wrong

I have frequently heard people cite John Ioannidis' apparent claim that "90% of medical research is false".

I think John Ioannidis is a brilliant person and I love his work and I think this statement points at a correct and important insight. But as phrased, I think this particular formulation when not paired with any caveats creates just a *little* more panic than is warranted.

Before I go further, Ioannidis' evidence:

He starts with simple statistics. Most studies are judged to have "discovered" a result if they reach $p < 0.05$, that is, if there is 5% probability or less the findings are due to mere chance (this is the best case scenario, where the study is totally free from bias or methodological flaws).

Suppose you throw a dart at the Big Chart O' Human Metabolic Pathways and supplement your experimental group with the chemical you hit. Then ten years later you come back and see how many of them died of heart attacks.

Most chemicals on the Big Chart probably don't prevent heart attacks. Let's say only one in a thousand do. Maybe your study will successfully find that 1/1000. But the 999 inactive chemicals will also throw up about 50 (999 * 5%) false positives significant at the 5% level. Therefore, even if you conduct your study perfectly, and it shows a significant decrease in heart attacks, there's about a 98% chance it's false.

One would hope medical scientists plan their studies with a little more care than throwing a dart at a metabolic chart. Yet many don't; a lot of genetic research is conducted by checking

every single gene against the characteristic of interest and seeing if any stick. And even when scientists have well-thought out theories, the inherent difficulty of medicine means they probably have less than a 50-50 chance of being right the first time, which means a 5% significance level has a less than 5% predictive value.

And this isn't even counting publication bias or poor methodology or conflicts of interest or anything like that.

Disturbingly, this problem seems to be borne out in empirical tests. Amgen Pharmaceuticals says it repeated experiments in 53 important papers and was only able to confirm 6. And Ioannidis himself did a re-analysis which is quoted as finding that "41% of the most influential studies in medicine have been convincingly shown to be wrong or significantly exaggerated."

So I don't at all disagree with the general consensus that this is a huge problem. But I do disagree with the following statements:

1. 90% of all medical research is wrong
2. A given study you read, or your doctor reads, is 90% likely to be wrong.
3. 90% of the things doctors believe, presumably based on these medical findings, is wrong.
4. This proves the medical establishment is clueless and hopelessly irrational and that two smart people working in a basement for five minutes can discover a new medical science far better than what all doctors could have produced in seventy years.

Is 90% of all medical research wrong?

As far as I can tell, there is no source at all for the 90% figure. I can't find it in any of Ioannidis' studies and indeed they

contradict it. His table of predictive values of different studies doesn't have *any* entries that correspond to 90% ("underpowered exploratory epidemiological study" is relatively close with 88%, but this is just for that one type of study, which is known to be especially bad). *The Atlantic* sums it up as:

> His model predicted, in different fields of medical research, rates of wrongness roughly corresponding to the observed rates at which findings were later convincingly refuted: 80 percent of non-randomized studies (by far the most common type) turn out to be wrong, as do 25 percent of supposedly gold-standard randomized trials, and as much as 10 percent of the platinum-standard large randomized trials.

Notice which number is conspicuously missing from that excerpt.

Now another study of his did show that in 90% of studies with very large effect sizes, later research eventually found the effect size to be smaller, but this was out of a pool of studies *specifically selected* for being surprising and likely to be false. I don't think it's the source of the number and if it were that would be terrible.

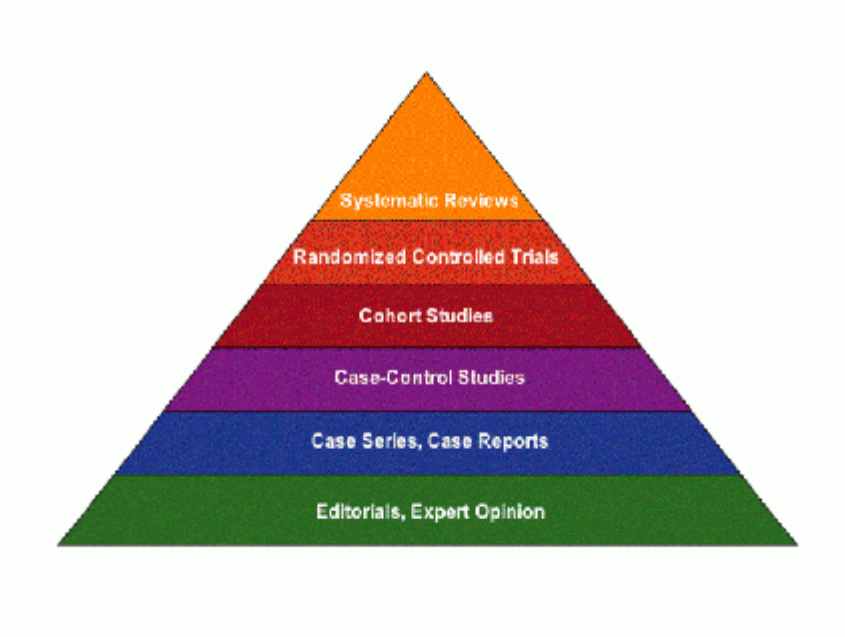
As far as I can tell, this started from a quote in an *Atlantic* article on Ioannidis which included the line "he charges that as much as 90 percent of the published medical information that doctors rely on is flawed". This then got turned into the title of a *Time* article "A Researcher's Claim: 90% of Medical Research Is Wrong", which itself got perverted to 90% of Medical Research Is Completely False.

So an unsourced quote that *up to* 90% of studies are *flawed* has somehow turned into a rallying cry that it has been *proven* that *at least* 90% of studies are *false*. To take this seriously we would have to believe that the numbers for *all research* are the same as the numbers for the poorly conducted epidemiological studies or the studies specifically selected for surprising results. I guess having a nice round number is good insofar as it makes the public pay attention to this field, but as far as actual numbers go, it's kind of made up.

Is any given study you read, or your doctor reads, 90% likely to be wrong?

But let's take the above number at face value and say that 90% of medical studies are wrong. Fine. Does that mean the last medical study you read about in Scientific American, or that your doctor used to recommend you a new drug, is wrong?

No. Let's look at the Medical Evidence Pyramid.



The medical evidence pyramid is much like all pyramids, in that the bottom levels are infested with snakes and booby traps and vengeful medical evidence mummies. It's only after you

reach the top few levels that you get the gold and jewels and precious, precious mummy powder.

This plays out in the same table of Ioannidis' speculations we saw before. While an *in vitro* study of the type used to identify possible drug targets might have a positive predictive value of 0.1%, a good meta-analysis or great RCT has a positive predictive value of 85%; that is, it's 85% likely to be true.

There are only two reasons someone might hear about the studies on the snake-infested bottom levels of the pyramid. Number one, that person is a specialist in the field who is valiantly trying to read through the entire niche medical journal the paper was published in. Or number two, the study found something incredible like DONUTS CURE CANCER IN A SAMPLE OF THREE LAB RATS!!! and the media decided to pick up on it. Hopefully everyone already ignores studies of the DONUTS CURE CANCER IN A SAMPLE OF THREE LAB RATES!!! type studies; if not, there's really not much I can say to you.

But *most* of the medical results that you hear about are the ones that get published in important journals and are trumpeted far and wide as important medical results. These are closer to the top of the pyramid than to the bottom. They're usually big expensive studies on thousands of people. Since the universities, hospitals, and corporations sponsoring them aren't idiots, they usually hire a decent statistician or two to make sure that they don't spend $300,000 testing something only to have a letter to the editor of the NEJM point out that they forgot to blind their subjects so it's totally worthless. And finally, in many cases you would only run a study that big and expensive if you had something plausible to test – you're not going to spend $300,000 just to throw a dart at the Big Chart O' Human Metabolic Pathways and see what happens.

So these studies that people actually hear about are bigger, they have more incentives to get their methodology right, and they're testing propositions with high plausibility. How do they do?

I said above that one of Ioannidis' studies was frequently quoted as saying that "41% of the most influential studies in medicine have been convincingly shown to be wrong or significantly exaggerated."

This is from a great study I totally endorse, but the 41% number was maximized for scariness. If I wanted to bias my reporting the other direction, I could equally well report the same results as "Only about 5% of influential medical experiments with adequate sample size have later been contradicted."

How? Ioannidis got his result by taking all medical studies with over 1000 citations in the '90s, of which there were 49. Of these, 4 were negative results (ie "X doesn't work") so he threw them out. This is the first part I think is kind of unfair. Yes, negative results aren't as sexy as positive results, but they're still influential medical research, and if Ioannidis is quoted as saying that X% of medical findings are later contradicted when he means that X% of positive medical findings are, that's not quite fair.

Annnnyway, of the 45 famous studies with positive findings, 11 didn't really get tested and so we don't know if they're right or wrong. Eliminating these is *also* a potential bias, because we expect that studies which seem sketchy are more likely to be replicated so people can find out if they're actually right. Ioannidis quite rightly set himself a higher bar by not eliminating them, but the quote about 41% of studies being wrong *does* seem to have gone back eliminated them – at least

that's the only way I can make the study numbers add up to 41% (the numbers given in the study actually say 32% of these studies failed to replicate).

So our 41% number is based off of 34 studies, best described as "34 famous medical studies that found positive findings ie the least believable kind of finding, plus were suspicious enough that someone wanted to replicate them".

Of these 34 studies, 7 were outright contradicted. Bad? Definitely. But for example, one of them was a study with a sample size of nine patients. Another study may well have been correct, but the results were interpreted wrongly (it said that estrogen decreased lipoprotein levels which everyone assumed meant decreased heart disease, but in fact later studies found increased heart disease without necessarily disproving the lipoprotein levels). Five of the six others were epidemiological trials, firmly on the middle of the pyramid. Only two of these contradicted studies were a true experiment with a sample size of >10.

(even here, I am sort of skeptical. Three of these disproven studies, two epidemiologicals and an experimental, purported to show Vitamin E decreased heart disease. Then a single better trial showed that Vitamin E did not decrease heart disease. While recognizing the last trial was better, it does seem like something more complicated is going on here than "all three of the earlier trials were just wrong", and I've recently been convinced antioxidant research is a huge minefield where tiny differences in protocol can cause big differences in results. But fine, let's grant this one and say there were two outright-contradicted experiments.)

So aside from the seven that were outright wrong, another seven were listed as "overstating their results".

There are a couple of problems that bothered me here. One of them was that Ioannidis decided to count studies as contradicting each other if relative risk in one study was half or less than in the other study, "regardless of whether confidence intervals might overlap or not". So even if a study effectively said "Here is a wide range of possible results, we think it's about here in the middle but our research is consistent with it being anywhere in this range", if another study got somewhere else in that range, the first study was marked as "exaggerated".

The second problem is, once again, poor studies versus poor interpretations. Ioannidis cites as an example of an exaggerated study one lasting a year and showing that the drug zidovudine helped slow the progression of HIV to AIDS. It concluded that giving HIV patients long-term zidovudine was probably a good idea. A later study lasted longer, and said that yes, zidovudine worked for a year, but then it stopped working. Because the earlier study had suggested longer-term zidovudine, it was marked as "exaggerated results", even though the results of both studies were totally consistent with one another (both found that zidovudine worked for the first year). This is probably of little consolation to AIDS patients who were treated with a useless drug, but it seems pretty important if we're investigating study methodology.

So the way I got my 5% figure was to take the two experimental studies with decent sample sizes which were actually contradicted and compare them to the 38 large experimental studies total that started the experiment.

So this suggests that if you see a large experimental study being trumpeted in the medical literature, the chance that it will be found to be totally false (as opposed to true but exaggerated) within ten years or so is only about 5% – which

if you understand p-values is about what you should have believed already.

(I think. This requires quite a few assumptions, not the least of which is that my calculations above are correct!)

Also worth noting: Ioannidis' experiment did not investigate the absolute highest level of the medical pyramid, systematic reviews and meta-analyses. I expect the best of these to be better than any individual study.

3. Are 90% of the things doctors believe, presumably based on medical findings, wrong?

After going through the steps above, it should be pretty obvious that the answer is no, because doctors are mostly reading famous influential studies like the ones mentioned above, which are at worst 40% and at best 5% wrong.

But there's another factor to be taken into account, which is that *why would you only read one study on something when lots of important findings have been investigated multiple times?*

Suppose that you're throwing darts at the Big Chart O' Human Metabolic Pathways, with your 1/1000 base rate of true hypotheses. You run a very good methodologically sound study and find p = .05. But now there's still only a 1/50 chance your hypothesis is correct.

But another team in China runs the same study, and they *also* find p = .05. We expect the Chinese to get false to true results at a rate of one to two (because the 1 in the 1/50 stays 1, but the 50 is divided by 20 to produce approximately 2. Wow, I'm even worse at explaining math than I am at doing it.)

Now a team in, oh, let's say Turkey runs the same study, and they *also* find p = .05. We expect the Turks to get false to true

results at a rate of one to ten, for, uh, the same math reasons as the Chinese. When the, um, Icelanders repeat the study, our odds go to one to two hundred.

So we started with 1000:1 odds, the first study brought us up to 50:1 odds, the second study to 2:1 odds, the third study to 1:10 odds, and the fourth study to 1:200 odds, ie we are now 99.5% sure we're right.

Real medicine is both better and worse than this. It's better in that we often have dozens of studies rather than just four. It's worse in that the studies are not all so methodologically sound that we can multiply our odds by 20 each time (to put it lightly).

But some of them are, and once we get enough of them, the base rate problems which plague individual medical findings go away very quickly. Even if only one of the studies is methodologically sound, if the reason they're studying their topic is because a bunch of other less believable studies all got positive results, that's a much better base rate than "because I hit it with my dart".

When doctors say that, for example, iron supplements help anaemia, it's not because they hit iron on their Big Chart O' Human Metabolic Pathways, then ran a single study, got p = .05, and rushed off to publish a medical textbook. It's because they knew hemoglobin had iron in it, there are at least 21 randomized controlled studies, probably some had p-values closer to .001 than to .05 even though I don't have any of them in front of me to check, and eventually some really really smart statisticians at the Cochrane Collaboration gave it their seal of approval. Most doctors' beliefs aren't on quite this high a level, but most doctors' beliefs aren't on the "Someone threw a dart, then did one study" level either.

4. Does this prove the medical establishment is clueless and hopelessly irrational and that two smart people working in a basement for five minutes can discover a new medical science far better than what all doctors could have produced in seventy years?

A lot of people seem to go from Ioannidis' experiment to something like "So I guess everyone in medicine is just clueless about how science and statistics work. I'll go read a couple of medical studies and then be able to outperform everyone in this totally flawed field."

(important note: I'm *not* accusing MetaMed of this! They seem pretty sane. I am accusing some people I come across in the community who are much more enthusiastic than the relatively sober MetaMed people of doing something like this.)

But the problem isn't that no one in medicine is familiar with Ioannidis' research. It's that they're not really sure what to do about it and figuring out a plan and implementing it will take time and effort.

Ioannidis' work isn't exactly secret. I've hung out with groups of residents (ie trainee doctors) who have discussed Ioannidis' findings over the dinner table. According to *The Atlantic*

> To say that Ioannidis's work has been embraced would be an understatement. His PLoS Medicine paper is the most downloaded in the journal's history, and it's not even Ioannidis's most-cited work—that would be a paper he published in Nature Genetics on the problems with gene-link studies. Other researchers are eager to work with him: he has published papers with 1,328 different co-authors at 538 institutions in 43 countries, he says. Last year he received, by his estimate, invitations to speak at

> 1,000 conferences and institutions around the world, and he was accepting an average of about five invitations a month until a case last year of excessive-travel-induced vertigo led him to cut back.

So if so many people are aware of this, why isn't the problem getting fixed more quickly?

An optimist could say the problem isn't getting fixed because there is no problem. A vast volume of embarassingly wrong medical literature gets published, inflates the publishers' resumes, and everyone else ignores it and concentrates on the not-really-so-bad large randomized trials. To the post-cynic it is all a smooth, well-functioning machine.

A pessimist might say that the problem isn't getting fixed because it's impossible. The average medical hypothesis is always going to have a low base rate of being true – in fact, if we force scientists to only study high base-rate hypotheses, by definition everything we discover will be boring. There will never be enough resources to apply huge rigorous trials to every one of the millions of things worth studying. So we're always going to have weak studies about low-base rate hypotheses, which is what Ioannidis is attacking as the recipe for failure.

A realist might point out there are some things we can do, but it involves coordinating a huge and complicated system with many moving parts. Journals can force trials to register before they conduct their experiments to avoid publication bias. The scientific community can give more status to people who perform important replications and especially important negative replications. Study authors and the media can come up with better ways to report their results to doctors and the public without blowing them out of proportion. Statisticians

can…actually, anything I say statisticians can do is just going to be a mysterious answer, along the lines of "do better statistics stuff", so I'm not going to embarass myself by completing this sentence except to postulate that I'll *bet* there's some recommendation that could complete it usefully.

But all these things involve vague entities who aren't really actors ("the scientific community", "the media") acting in ways that are kind of against their immediate incentives. This is hard to make people do and usually involves a lot of grassroots coordination effort. Which is going on. But it takes time.

But no matter what happens, I think a useful epistemic habit is to be very skeptical of individual studies, and skeptical but not *too* skeptical of large randomized trials, good meta-analyses, and general medical consensus when supported by an evidence base.

Prisons are Built with Bricks of Law and Brothels with Bricks of Religion, But That Doesn't Prove a Causal Relationship

Research Suggests Psychiatric Interventions Like Admission To A Mental Hospital Could Increase Suicide Risk says an Alternet article about a study that specifically mentions that it should not be used to conclude that psychiatric interventions like admission to a mental hospital could increase suicide risk.

But I wouldn't be so worried if it wasn't based on a very similar editorial written by field experts and published in the Journal of Social Psychiatry and Psychiatric Epidemiology.

The study involved is Rygaard-Hjorthøj, Madsen, Agerbo, and Nordentoft (2013), hereafter just "Hjorthøj" because I like saying that word. Hjorthøj finds that people who receive psychiatric treatment are much more likely to commit suicide than people who don't. For example, someone who gets psychiatric medication is six times more likely to commit suicide than someone who doesn't; someone who gets admitted to a psychiatric hospital is a whopping 44 times more likely to commit suicide than someone who doesn't. The authors observe a "dose-response relationship", which means that the more psychiatric treatment you get, the more likely you are to kill yourself.

Now, you're probably asking yourself at this point "Wait, were they just using perfectly healthy people with no psychiatric problems as a control group?" and the answer is yes. Yes they were. So this study is *basically* finding that people who get committed to psychiatric hospitals are more likely to be the

sort of people who are going to commit suicide than people who do not get committed to psychiatric hospitals. I for one find this result rather reassuring.

The authors of the study are absolutely on board with this, saying that "observational studies such as the present one cannot establish causality, but merely associations", and their conclusion is that "not only people with a history of of psychiatric hospitalization, but also those receiving only psychiatric medication, outpatient treatment, or emergency room treatment should be monitored more closely". Sure. If you absolutely must have a snappier conclusion than "psych patients often mentally ill, more at eleven," I guess that fits the bill.

But according to an editorial published in the same journal by two people who are *not* the original authors, it says something much more sinister:

> The results of a study in this issue of the Journal…raise the disturbing possibility that psychiatric care might, at least in part, cause suicide.

A…bold hypothesis. Why should we privilege this hypothesis over the alternative possibility that suicidal people are more likely to seek (or get forced into) psychiatric treatment?

> The authors understandably caution that 'the association is likely one of selection rather than causation, in that people with increasing levels of psychiatric contract are also more severely at risk of dying from suicide.' This is undoubtedly part of the reason for the association, but it is not possible to be sure that an element of causation may not also be contributing. Associations that are strong, demonstrate a dose-effect relationship, and have a

> plausible mechanism are more likely to indicate a causal relationship than associations that lack these characteristics.

And then the Alternet article picks this up and adds a different argument:

> The Danish researchers argued that we were seeing the results of something like a cancer treatment study. Sicker people were appropriately getting into more intensive treatments, but unfortunately the sicker they were the more likely it was that they would still die, despite even the best of medicines. They also suggested that we may have therefore discovered the most accurate predictor of suicide we've ever found: The more someone seeks or is forced into psychiatric care, the closer they probably are on the trajectory towards suicide.
>
> The only problem with this line of reasoning is that there's no evidence to support it. Suicide is not a progressive illness like cancer; that is, there's no evidence that people with suicidal feelings travel on a trajectory of ever-intensifying, ever-more-constant suicidal feelings while getting into ever more intensive psychiatric care until they die at steadily increasing rates along the way. If suicidality was in fact progressive in that way, we'd be much better at identifying where people are along that path and intervening at the right time to prevent suicides. Instead, completed suicides tend to be impulsive, related to a myriad of cascading, confounding, unpredictable factors, not much more common overall in people diagnosed with mental disorders than in the general population, and most often surprising to even those closest to the victims.

Okay, let's stop talking about psychiatric disease and shift to murder.

Probably the best risk factor for murder that you will ever find, better than being abused as a child or doing drugs or having the MAOA warrior gene or whatever, is "previous contact with the police".

Murder is not "progressive" (shut up, neoreactionaries). Much like suicide, there's no evidence that murderers "travel on a trajectory of ever-intensifying, ever-more-constant murderous feelings while getting into more intensive police custody until they kill at steadily increasing rates along the way." Instead it seems to be "impulsive, related to a myriad of cascading, confounding, unpredictable factors, and surprising even to those closest to the perpetrators."

The link between murder and previous contact with the police will be strong. For example, previous murderers released from prison have a 1.2% chance of getting arrested for another murder within three years, compared to about a 0.0001% murder rate per three years among the general population. That's a relative risk of 10,000x, which blows Hjorthøj's relative risk of 44x out of the water.

The link will be dose-dependent. People who have previously only gotten warnings from the police will be less likely to murder than people who have gotten small fines, who are less likely to murder than people who have gotten probation, who are less likely to murder than people who have gotten short jail sentences, who are less likely to murder than people who have gotten long jail sentences.

The link even has a plausible causal mechanism. Contact with the police can seriously disrupt people's lives, making them stressed and anxious and angry and hopeless, all of which are

the sort of emotions that predispose someone towards violence.

Therefore, the police cause murder?

Here are some other links that are non-progressive, strong, dose-dependent, and have plausible causal mechanisms.

The link between getting detention and dropping out of school. Therefore, detentions cause students to become demoralized and drop out from school.

The link between ice cream sales in a city and heatstroke cases in that city. Therefore, ice cream contains toxic chemicals that cause heatstroke.

The link between having lots of bruises and being in an abusive relationship. Therefore, abusers only abuse their victims because they're angry about how many bruises they have.

The editorial authors seem to have gotten the "strong, dose-dependent, plausible" criteria from an article on epidemiology (God only knows where the journalist got the non-progressive criterion from). I would bet that the epidemiology article either did not intend for it to be used in this way, or that it meant that these criteria provide only the most tenuous of possible links.

This is why the saying is "correlation doesn't imply causation" and not "correlation does not imply causation, unless it's really *strong* correlation, in which case knock yourself out."

And this is why the article finds that even going to a psychiatric emergency room and being turned down for treatment increases your risk of suicide almost twenty times. I mean, in *my* ER patients only even see a psychiatrist for like half an hour. You're saying a half an hour with a psychiatrist

leads to a vigintupling of suicide rates months down the road? We might be bad. But we're not *that* bad.

The sad thing is, I think there might be a point buried underneath all this.

You can't conclude from an increased murder rate among people with criminal histories that the police cause murder. But the justice system *does* contribute to murder in its way by sticking hardened criminals together, traumatizing them, and failing to give them enough resources to rebuild their lives. The contribution of the criminal justice system to crime isn't exactly a secret, it's just not accessible with that methodology.

Likewise, I don't disagree that contact with the psychiatric system can sometimes be harmful. Forced commitment can sometimes make people lose their jobs, or cause them stigma, or stick them in an unpleasant psychiatric hospital where they don't want to be. While there are no doubt potential benefits as well, the weighing of the costs and benefits is something that hasn't been investigated nearly as much as it deserves. I think forced committment is an overused tool and would be glad to get some evidence backing me up.

But this paper contributes *nothing* to the discussion. All we know is there's an association between psychiatric care and suicide, which was entirely obvious already. We don't know how much of that association is causal, how much of it is selection, and how much of it is "it would be even worse without psychiatric care but psychiatric care can't do *everything*.

The exact effect of psychiatric care on suicide is a topic worthy of further high-quality research and discussion. But this isn't it.

Noisy Poll Results and the Reptilian Muslim Climatologists from Mars

Beware of Phantom Lizardmen

I have only done a little bit of social science research, but it was enough to make me hate people. One study I helped with analyzed whether people from different countries had different answers on a certain psychological test. So we put up a website where people answered some questions about themselves (like "what country are you from?") and then took the psychological test.

And so of course people screwed it up in every conceivable way. There were the merely dumb, like the guy who put "male" as his nationality and "American" as his gender. But there were also the actively malicious or at least annoying, like the people (yes, more than one) who wrote in "Martian".

I think we all probably know someone like this, maybe a couple people like this.

I also think most of us *don't* know someone who believes reptilian aliens in human form control all the major nations of Earth.

Public Policy Polling's recent poll on conspiracy theories mostly showed up on my Facebook feed as "Four percent of Americans believe lizardmen are running the Earth".

(of note, an additional 7% of Americans are "not sure" whether lizardmen are running the Earth or not.)

Imagine the situation. You're at home, eating dinner. You get a call from someone who says "Hello, this is Public Policy Polling. Would you mind answering some questions for us?"

You say “Sure”. An extremely dignified sounding voice says – and this is the exact wording of the question – “Do you believe that shape-shifting reptilian people control our world by taking on human form and gaining political power to manipulate our society, or not?” Then it urges you to press 1 if yes, press 2 if no, press 3 if not sure.

So first we get the people who think “Wait, was 1 the one for if I did believe in lizardmen, or if I didn’t? I’ll just press 1 and move on to the next question.”

Then we get the people who are like “I never heard it before, but if this nice pollster thinks it’s true, I might as well go along with them.”

Then we get the people who are all “F#&k you, polling company, I don’t want people calling me when I’m at dinner. You screw with me, I tell you what I’m going to do. I’m going to tell you I believe lizard people are running the planet.”

And *then* we get the people who put “Martian” as their nationality in psychology experiments. Because some men just want to watch the world burn.

Do these three groups total 4% of the US population? Seems plausible.

I really wish polls like these would include a control question, something utterly implausible even by lizard-people standards, something like “Do you believe Barack Obama is a hippopotamus?” Whatever percent of people answer yes to the hippo question get subtracted out from the other questions.

Poll Answers As Attire

Alas, not all weird poll answers can be explained that easily. On the same poll, 13% of Americans claimed to believe Barack Obama was the Anti-Christ. Subtracting our

Lizardman's Constant of 4%, that leaves 9% of Americans who apparently gave this answer with something approaching sincerity.

(a friend on Facebook pointed out that 5% of *Obama voters* claimed to believe that Obama was the Anti-Christ, which seems to be another piece of evidence in favor of a Lizardman's Constant of 4-5%. On the other hand, I do enjoy picturing someone standing in a voting booth, thinking to themselves "Well, on the one hand, Obama is the Anti-Christ. On the other, do I really want four years of Romney?")

Some pollsters are starting to consider these sorts of things symptomatic of what they term symbolic belief, which seems to be kind of what the Less Wrong sequences call Professing and Cheering or Belief As Attire. Basically, people are being emotivists rather than realists about belief. "Obama is the Anti-Christ" is another way of just saying "Boo Obama!", rather than expressing some sort of proposition about the world.

And the same is true of "Obama is a Muslim" or "Obama was not born in America".

Never Attribute To Stupidity What Can Be Adequately Explained By Malice

But sometimes it's not some abstruse subtle bias. Sometimes it's not a good-natured joke. Sometimes people might just be actively working to corrupt your data.

Another link I've seen on my Facebook wall a few times is this one: Are Climate Change Sceptics More Likely To Be Conspiracy Theorists? It's based on a paper by Stephen Lewandowsky et al called NASA Faked The Moon Landing, Therefore Climate Science Is A Hoax – An Analysis Of The Motivated Rejection Of Science.

The paper's thesis was that climate change skeptics are motivated by conspiracy ideation – a belief that there are large groups of sinister people out to deceive them. This seems sort of reasonable on the face of it – being a climate change skeptic requires going against the belief of the entire scientific establishment. My guess is that there probably is a significant link here waiting to be discovered.

Unfortunately, it's…possible Stephan Lewandowsky wasn't the best person to investigate this? Aside from being a professor of cognitive science, he also runs Shaping Tomorrow's World, a group that promotes "re-examining some of the assumptions we make about our technological, social and economic systems" and which seems to be largely about promoting global warming activism. While I think it's admirable that he is involved in that, it raises conflict of interest questions. And the way his paper is written – starting with the over-the-top title – doesn't do him any favors.

(if the conflict of interest angle doesn't make immediate and obvious sense to you, imagine how sketchy it would be if a professional global warming *denier* was involved in researching the motivations of global warming *supporters*)

But enough of my personal opinions. What's the paper look like?

The methodology goes like this: they send requests to several popular climate blogs, both believer and skeptic, asking them to link their readers to an online survey. The survey asks people their beliefs on global warming and on lots of conspiracy theories and fringe beliefs.

On first glance, the results are extremely damning. People who rejected climate science were wildly more likely to reject pretty much every other form of science as well, including the

"theory" that HIV causes AIDS and the "theory" that cigarettes cause cancer. They were more willing to believe aliens landed at Roswell, that 9-11 was an inside job, and, yes, that NASA faked the moon landing. The conclusion: climate skeptics are just really stupid people.

But a bunch of global warming skeptics started re-analyzing the data and coming up with their own interpretations. They found that many large pro-global-warming blogs posted the link to the survey, but very few anti-global-warming blogs did. This then devolved into literally the worst flame war I have ever seen on the Internet, centering around accusations about whether the study authors deliberately excluded large anti-global warming blogs, or whether the authors asked the writers of anti-global-warming blogs and these writers just ignored the request (my impression is that most people now agree it was the latter). In either case, it ended up with most people taking the survey being from the pro-global-warming blogs, and only a few skeptics.

More interestingly, they found that pretty much all of the link between global warming skepticism and stupidity was a couple of people (there were so few skeptics, *and* so few conspiracy believers, that these couple of people made up a pretty big proportion of them, and way more than enough to get a "significant" difference with the global warming believers). Further, most of these couple of people had given the maximally skeptical answer to every single question about global warming, and the maximally credulous answer to every single question about conspiracies.

The danger here now seems obvious. Global warming believer blogs publish a link to this study, saying gleefully that it's going to prove that global warming skeptics are idiots who also think NASA faked the moon landing and the world is run

by lizardmen or whatever. Some global warming believers decide to help this process along by pretending to be super-strong global warming skeptics and filling in the stupidest answers they can to every question. The few real global warming skeptics who take the survey aren't enough signal to completely drown out this noise. Therefore, they do the statistics and triumphantly announce that global warming skepticism is linked to stupid beliefs.

The global warming skeptic blogosphere has in my opinion done more than enough work to present a very very strong case that this is what happened (somebody else do an independent look at the controversy and double-check this for me?) And Professor Lewandowsky's answer was…

…to publish a second paper, saying his results had been confirmed because climate skeptics were so obsessed with conspiracy theories that they had accused his data proving they were obsessed with conspiracies of being part of a conspiracy. The name of the paper? Recursive Fury. I have to hand it to him, this is possibly *the most chutzpah I have ever seen a single human being display.*

(the paper is now partially offline as the journal investigates it for ethical something something)

The lesson from all three of the cases in this post seems clear. When we're talking about very unpopular beliefs, polls can only give a weak signal. Any possible source of noise – jokesters, cognitive biases, or deliberate misbehavior – can easily overwhelm the signal. Therefore, polls that rely on detecting very weak signals should be taken with a grain of salt.

Two Dark Side Statistics Papers

I.

First we have False Positive Psychology: Undisclosed Flexibility In Data Collection And Analysis Allows Presenting Anything As Significant (h/t Jonas Vollmer).

The message is hardly unique: there are lots of tricks unscrupulous or desperate scientists can use to artificially nudge results to the 5% significance level. The clarity of the presentation *is* unique. They start by discussing four particular tricks:

1. Measure multiple dependent variables, then report the ones that are significant. For example, if you're measuring whether treatment for a certain psychiatric disorder improves life outcomes, you can collect five different measures of life outcomes – let's say educational attainment, income, self-reported happiness, whether or not ever arrested, whether or not in romantic relationship – and have a 25%-ish probability one of them will come out at significance by chance. Then you can publish a paper called "Psychiatric Treatment Found To Increase Educational Attainment" without ever mentioning the four negative tests.

2. Artificially choose when to end your experiment. Suppose you want to prove that yelling at a coin makes it more likely to come up tails. You yell at a coin and flip it. It comes up heads. You try again. It comes up tails. You try again. It comes up heads. You try again. It comes up tails. You try again. It comes up tails again. You try again. It comes up tails again. You note that it came up tails four out of six times – a 66% success rate compared to expected 50% – and declare victory. Of course, this result wouldn't be significant, and it seems as if this should be a general rule – that almost by the definition of significance, you shouldn't be able to obtain it just be stopping the experiment at the right point. But the authors of the study perform several simulations to prove that this trick is more successful than you'd think:

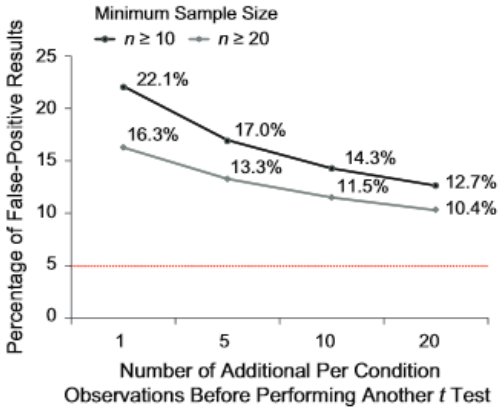


Fig. 1. Likelihood of obtaining a false-positive result when data collection ends upon obtaining significance ($p \leq .05$, highlighted by the dotted line). The figure depicts likelihoods for two minimum sample sizes, as a function of the frequency with which significance tests are performed.

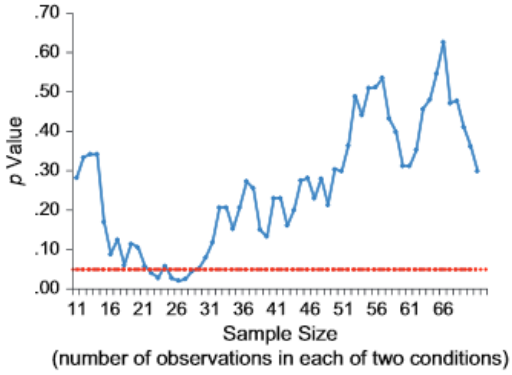


Fig. 2. Illustrative simulation of p values obtained by a researcher who continuously adds an observation to each of two conditions, conducting a t test after each addition. The dotted line highlights the conventional significance criterion of $p \leq .05$.

3. Control for "confounders" (in practice, most often gender). I sometimes call this the "Elderly Hispanic Woman Effect" after drug trials that find that their drug doesn't have significant effects in the general population, but it *does* significantly help elderly Hispanic women. The trick is you split the population into twenty subgroups (young white men, young white women, elderly white men, elderly white women, young black men, etc), in one of those subgroups it will achieve significance by pure chance, and so you declare that your drug must just somehow be a perfect fit for elderly Hispanic women's unique body chemistry. This is not *always* wrong (some antihypertensives have notably different efficacy in white versus black populations) but it is *usually* suspicious.

4. Test different conditions and report the ones you like. For example, suppose you are testing whether vegetable consumption affects depression. You conduct the trial with three arms: low veggie diet, medium veggie diet, and high veggie diet. You now have four possible comparisons – low-medium, low-high, medium-high, low-medium-high trend). One of them will be significant 20% of the time, so you can just report that one: "People who eat a moderate amount of vegetables are less likely to get depression than people who eat excess vegetables" sounds like a perfectly reasonable result.

Then they run simulations to show exactly how much more likely you are to get a significant result in random data by employing each

trick:

Table 1. Likelihood of Obtaining a False-Positive Result

	Significance level		
Researcher degrees of freedom	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

The image demonstrates that by using all four tricks, you can squeeze random data into a result significant at the $p < 0.05$ level about 61% of the time.

The authors then put their money where their mouth is by conducting two studies.

The first seems like a very very classic social psychology study. Subjects are randomly assigned to listen to one of two songs - either a nondescript control song or a child's nursery song. Then they are asked to rate how old they feel. Sure enough, the subjects who listen to the child's song feel older ($p = 0.03$).

The second study is very similar, with one important exception. Once again, subjects are randomly assigned to listen to one of two songs - either a nondescript control song or a song about aging - "When I'm Sixty-Four" by The Beatles. Then they are asked to put down their actual age, in years. People who listened to the Beatles song became, on average, a year and a half younger than the control group ($p = 0.04$).

So either the experimental intervention changed their subjects' ages, or the researchers were using statistical tricks. Turns out it was the second one. They explain how they used the four statistical tricks they explained above, and that without those tricks there would have been (obviously) no significant difference. They go on to say that their experiment meets the inclusion criteria for every major journal

and that under current reporting rules there's no way anyone could have detected their data manipulation.

They go on to list the changes they think the scientific establishment needs to prevent papers like theirs from reaching print. They're basically "don't do the things we just talked about", but as far as I can tell they rely on the honor system. I think a broader meta-point is that on important studies scientists should have to submit their experimental protocol to a journal and get it accepted or rejected in advance so they can't change tactics mid-stream or drop data. This would also force journals to publish more negative results.

See also their interesting discussion of why they think "use Bayesian statistics" is a non-solution to the problem.

II.

Second we have How To Have A High Success Rate In Treatment: Advice For Evaluators Of Alcoholism Programs.

This study is very close to my heart, because I'm working on my hospital's Substance Abuse Team this month. Every day we go see patients struggling with alcoholism, heroin abuse, et cetera, and we offer them treatment at our hospital's intensive inpatient Chemical Dependency Unit. And every day, our patients say thanks but no thanks, they heard of a program affiliated with their local church that has a 60% success rate, or an 80% success rate, or in one especially rosy-eyed case a frickin' 97% success rate.

(meanwhile, *real* rehab programs still struggle to prove they have a success rate greater than placebo)

My attending assumes these programs are scum but didn't really have a good evidence base for the claim, so I decided to search Google Scholar to find out what was going on. I struck gold in this paper, which is framed as a sarcastic how-to guide for unscrupulous drug treatment program directors who want to inflate their success rates without *technically* lying.

By far the best way to do this is to choose your denominator carefully. For example, it seems fair to only include the people who

attended your full treatment program, not the people who dropped out on Day One or never showed up at all – you can hardly be blamed for that, right? So suppose that your treatment program is one month intensive in rehab followed by a series of weekly meetings continuing indefinitely. At the end of one year, you define successful treatment completers as "the people who are still going to these meetings now, at the end of the year". But in general, people who relapse into alcoholism are a whole lot less likely to continue attending their AA meetings than people who stay sober. So all you have to do is go up to people at your AA meeting, ask them if they're still on the wagon, and your one-year success rate looks really good.

Another way to hack your treatment population is to only accept the most promising candidates to begin with (it works for private schools and it can work for you). We know that middle-class, employed people with houses and families have a much better prognosis than lower-class unemployed homeless single people. Although someone would probably notice if you put up a sign saying "MIDDLE-CLASS EMPLOYED PEOPLE WITH HOUSES AND FAMILIES ONLY", a very practical option is to just charge a lot of money and let your client population select themselves. This is why for-profit private rehabs will have a higher success rate than public hospitals and government programs that deal with poor people.

Still another strategy is to follow the old proverb: "If at first you don't succeed, redefine success". "Abstinence" is such a harsh word. Why not "drinking in moderation"? This is a wonderful phrase, because you can just let the alcoholic involved determine the definition of moderation. A year after the program ends, you can send out little surveys saying "Remember when we told you God really wants you not to drink? You listened to us and are drinking in moderation now, right? Please check one: Y () N ()". Who's going to answer 'no' to that? Heck, some of the alcoholics I talk to say they're drinking in moderation *while they are in the emergency room for alcohol poisoning*.

If you can't handle "moderation", how about "drinking less than you were before the treatment program"? This takes advantage of regression to the mean – you're going to enter a rehab program at the worst period of your life, the time when your drinking finally spirals out of control. Just by coincidence, most other parts of your life will include less drinking than when you first came in to rehab, including the date a year after treatment when someone sends you a survey. Clearly rehab was a success!

And why wait a year? My attending and myself actually looked up what was going on with that one 97% success rate program our patient said he was going to. Here's what they do – it's a three month residential program where you live in a building just off the church and you're not allowed to go out except on group treatment activities. Obviously there is no alcohol allowed in the building and you are surrounded by very earnest counselors and fellow recovering addicts at all times. Then, *at the end of the three months, while you are still in the building*, they ask you whether you're drinking or not. You say no. Boom – 97% success rate.

One other tactic I have actually seen in studies and it *breaks my heart* is interval subdivision, which reminds me of some of the dirty tricks from the first study above. At five years' follow-up, you ask people "Did you drink during Year 1? Did you drink during Year 2? Did you drink during Year 3?…" and so on. Now you have five chances to find a significant difference between treatment and control groups. I have literally seen studies that say "Our rehab didn't have an immediate effect, but by Year 4 our patients were doing better than the controls." Meanwhile, in years 1, 2, 3, and 5, for all we know the controls were doing better than the patients.

But if all else fails, there's always the old standby of poor researchers everywhere – just don't include a control group at all. This table really speaks to me:

Table 1. *Examples of the disappointing effect of control groups*

Study	Results for treatment group	Type of comparison group included	Results for comparison group
Ditman *et al.* (1967)	32% success rate (no new arrests) among clinic-treated offenders at 12 months	Probation alone, without treatment	44% success rate (no new arrests)
Fuller *et al.* (1986)	18.8% continuously abstinent for 1 year with disulfiram	Placebo medication	22.5% continuously abstinent for 1 year
Miller *et al.* (1981)	80% of therapist-treated cases improved at 6 months	Self-help manual and minimal therapist contact	87% of cases improved at 6 months
Sanchez-Craig *et al.* (1991)	71% of therapist-treated cases problem-free at 12 months	Self-help materials and brief therapist contact	71% of cases problem-free at 12 months

The great thing about this table isn't just that it shows that seemingly impressive results are exactly the same as placebo. The great thing it shows is that results in the placebo groups in the four studies could be anywhere from a 22.5% success rate to an 87% success rate. These aren't treatment differences – all four groups are placebo! This is one hundred percent a difference in study populations and in success measures used. In other words, depending on your study protocol, you can prove that there is a 22.5% chance the average untreated alcoholic will achieve remission, or an 87% chance the average untreated alcoholic will achieve remission.

You can bet that rehabs use the study protocol that finds an 87% chance of remission in the untreated. And then they go on to boast of their 90% success rate. Good job, rehab!

Alcoholics Anonymous: Much More Than You Wanted to Know

[EDIT 10/27: Slight changes in response to feedback; correcting some definitions. I am not an expert in this field and will continue to make changes as I learn about them. There is a critique of this post here and other worse critiques elsewhere. My only excuse for doing this is that I am failing less spectacularly than other online sources writing about the same topic.]

I've worked with doctors who think Alcoholics Anonymous is so important for the treatment of alcoholism that anyone who refuses to go at least three times a week is in denial about their problem and can't benefit from further treatment.

I've also worked with doctors who are so against the organization that they describe it as a "cult" and say that a physician who recommends it is no better than one who recommends crystal healing or dianetics.

I finally got so exasperated that I put on my Research Cap and started looking through the evidence base.

My conclusion, after several hours of study, is that now I understand why most people don't do this.

The studies surrounding Alcoholics Anonymous are some of the most convoluted, hilariously screwed-up research I have ever seen. They go wrong in ways I didn't even realize research *could* go wrong before. Just to give some examples:

– In several studies, subjects in the "not attending Alcoholics Anonymous" condition attended Alcoholics Anonymous more than subjects in the "attending Alcoholics Anonymous" condition.

– Almost everyone's belief about AA's retention rate is off by a factor of five because one person long ago misread a really confusing graph and everyone else copied them without double-checking.

– The largest study ever in the field, a $30 million effort over 8 years following thousands of patients, had no untreated control group.

Not only are the studies poor, but the people interpreting them are heavily politicized. The entire field of addiction medicine has gotten stuck in the middle of some of the most divisive issues in our culture, like whether addiction is a biological disease or a failure of willpower, whether problems should be solved by community and peer groups or by highly trained professionals, and whether there's a role for appealing to a higher power in any public organization. AA's supporters see it as a scruffy grassroots organization of real people willing to get their hands dirty, who can cure addicts failed time and time again by a system of glitzy rehabs run by arrogant doctors who think their medical degrees make them better than people who have personally fought their own battles. Opponents see it as this awful cult that doesn't provide any real treatment and just tells addicts that they're terrible people who will never get better unless they sacrifice their identity to the collective.

As a result, the few sparks of light the research kindles are ignored, taken out of context, or misinterpreted.

The entire situation is complicated by a bigger question. We will soon find that AA usually does not work better or worse than various other substance abuse interventions. That leaves the sort of question that all those fancy-shmancy people with control groups in their studies don't have to worry about – does anything work at all?

I.

We can start by just taking a big survey of people in Alcoholics Anonymous and seeing how they're doing. On the

one hand, we don't have a control group. On the other hand… well, there really is no other hand, but people keep doing it.

According to AA's own surveys, one-third of new members drop out by the end of their first month, half by the end of their third month, and three-quarters by the end of their first year. "Drop out" means they don't go to AA meetings anymore, which could be for any reason including (if we're feeling optimistic) them being so completely cured they no longer feel they need it.

There is an alternate reference going around that only 5% (rather than 25%) of AA members remain after their first year. This is a mistake caused by misinterpreting a graph showing that only five percent of members in their first year were in their twelfth month of membership, which is obviously completely different. Nevertheless, a large number of AA hate sites (and large rehabs!) cite the incorrect interpretation, for example the Orange Papers and RationalWiki's page on Alcoholics Anonymous. In fact, just to keep things short, assume RationalWiki's AA page makes every single mistake I warn against in the rest of this article, then use that to judge them in general. On the other hand, Wikipedia gets it right and I continue to encourage everyone to use it as one of the most reliable sources of medical information available to the public (I wish I was joking).

This retention information isn't very helpful, since people can remain in AA without successfully quitting drinking, and people may successfully quit drinking without being in AA. However, various different sources suggest that, of people who stay in AA a reasonable amount of time, about half stop being alcoholic. These numbers can change wildly depending on how you define "reasonable amount of time" and "stop being

alcoholic". Here is a table, which I have cited on this blog before and will probably cite again:

Table 1. *Examples of the disappointing effect of control groups*

Study	Results for treatment group	Type of comparison group included	Results for comparison group
Ditman *et al.* (1967)	32% success rate (no new arrests) among clinic-treated offenders at 12 months	Probation alone, without treatment	44% success rate (no new arrests)
Fuller *et al.* (1986)	18.8% continuously abstinent for 1 year with disulfiram	Placebo medication	22.5% continuously abstinent for 1 year
Miller *et al.* (1981)	80% of therapist-treated cases improved at 6 months	Self-help manual and minimal therapist contact	87% of cases improved at 6 months
Sanchez-Craig *et al.* (1991)	71% of therapist-treated cases problem-free at 12 months	Self-help materials and brief therapist contact	71% of cases problem-free at 12 months

Behold. Treatments that look very impressive (80% improved after six months!) turn out to be the same or worse as the control group. And comparing control group to control group, you can find that "no treatment" can appear to give wildly different outcomes (from 20% to 80% "recovery") depending on what population you're looking at and how you define "recovery".

Twenty years ago, it was extremely edgy and taboo for a reputable scientist to claim that alcoholics could recover on their own. This has given way to the current status quo, in which pretty much everyone in the field writes journal articles all the time about how alcoholics can recover on their own, but make sure to harp upon how edgy and taboo they are for doing so. From these sorts of articles, we learn that about 80% of recovered alcoholics have gotten better without treatment, and many of them are currently able to drink moderately without immediately relapsing (something *else* it used to be extremely taboo to mention). Kate recently shared an good article about this: Most People With Addiction Simply Grow Out Of It: Why Is This Widely Denied?

Anyway, all this stuff about not being able to compare different populations, and the possibility of spontaneous recovery, just mean that we need controlled experiments. The largest number of these take a group of alcoholics, follow them closely, and then evaluate all of them – the AA-attending and the non-AA-attending – according to the same criteria. For example Morgenstern et al (1997), Humphreys et al (1997) and Moos (2006). Emrick et al (1993) is a meta-analyses of *a hundred seventy three* of these. All of these find that the alcoholics who end up going to AA meetings are much more likely to get better than those who don't. So that's good evidence the group is effective, right?

Bzzzt! No! Wrong! Selection bias!

People who want to quit drinking are more likely to go to AA than people who don't want to quit drinking. People who want to quit drinking are more likely to *actually* quit drinking than those who don't want to. This is a *serious* problem. Imagine if it is common wisdom that AA is the best, maybe the only, way to quit drinking. Then 100% of people who really want to quit would attend compared to 0% of people who didn't want to quit. And suppose everyone who wants to quit succeeds, because secretly, quitting alcohol is really easy. Then 100% of AA members would quit, compared to 0% of non-members – the most striking result it is mathematically possible to have. And yet AA would not have made a smidgeon of difference.

But it's worse than this, because attending AA isn't just about wanting to quit. It's also about having the resources to make it to AA. That is, wealthier people are more likely to hear about AA (better information networks, more likely to go to doctor or counselor who can recommend) and more likely to be able to attend AA (better access to transportation, more flexible job schedules). But wealthier people are also known to be better at

quitting alcohol than poor people – either because the same positive personal qualities that helped them achieve success elsewhere help them in this battle as well, or just because they have fewer other stressors going on in their lives driving them to drink.

Finally, perseverance is a confounder. To go to AA, and to keep going for months and months, means you've got the willpower to drag yourself off the couch to do a potentially unpleasant thing. That's probably the same willpower that helps you stay away from the bar.

And then there's a confounder going the *opposite* direction. The worse your alcoholism is, the more likely you are to, as the organization itself puts it, "admit you have a problem".

These sorts of longitudinal studies are almost useless and the field has mostly moved away from them. Nevertheless, if you look on the pro-AA sites, you will find them in droves, and all of them "prove" the organization's effectiveness.

III.

It looks like we need randomized controlled trials. And we have them. Sort of.

Brandsma (1980) is the study beloved of the AA hate groups, since it purports to show that people in Alcoholics Anonymous not only don't get better, but are *nine times* more likely to binge drink than people who don't go into AA at all.

There are a number of problems with this conclusion. First of all, if you actually look at the study, this is one of about fifty different findings. The other findings are things like "88% of treated subjects reported a reduction in drinking, compared to 50% of the untreated control group".

Second of all, the increased binge drinking was significant at the 6 month followup period. It was *not* significant at the end of treatment, the 3 month followup period, the 9 month followup period, or the 12 month followup period. Remember, taking a single followup result out of the context of the other followup results is a classic piece of Dark Side Statistics and will send you to Science Hell.

Of multiple different endpoints, Alcoholics Anonymous did better than no treatment on almost all of them. It did worse than other treatments on some of them (dropout rates, binge drinking, MMPI scale) and the same as other treatments on others (abstinent days, total abstinence).

If you are pro-AA, you can say "Brandsma study proves AA works!". If you are anti-AA, you can say "Brandsma study proves AA works worse than other treatments!", although in practice most of these people prefer to quote extremely selective endpoints out of context.

However, most of the patients in the Brandsma study were people convicted of alcohol-related crimes ordered to attend treatment as part of their sentence. Advocates of AA make a good point that this population might be a bad fit for AA. They may not feel any personal motivation to treatment, which might be okay if you're going to listen to a psychologist do therapy with you, but fatal for a *self*-help group. Since the whole point of AA is being in a community of like-minded individuals, if you don't actually feel any personal connection to the project of quitting alcohol, it will just make you feel uncomfortable and out of place.

Also, uh, this just in, Brandsma didn't use a real AA group, because the real AA groups make people be anonymous which makes it inconvenient to research stuff. He just sort of started

his own non-anonymous group, let's call it A, with no help from the rest of the fellowship, and had it do Alcoholics Anonymous-like stuff. On the other hand, many members of his control group went out into the community and…attended a real Alcoholics Anonymous, because Brandsma can't exactly ethically tell them not to. So *technically*, there were more people in AA in the no-AA group than in the AA group. Without knowing more about Alcoholics Anonymous, I can't know whether this objection is valid and whether Brandsma's group did or didn't capture the essence of the organization. Still, not the sort of thing you want to hear about a study.

Walsh et al (1991) is a similar study with similar confounders and similar results. Workers in an industrial plant who were in trouble for coming in drunk were randomly assigned either to an inpatient treatment program or to Alcoholics Anonymous. After a year of followup, 60% of the inpatient-treated workers had stayed sober, but only 30% of the AA-treated workers had.

The pro-AA side made three objections to this study, of which one is bad and two are good.

The bad objection was that AA is cheaper than hospitalization, so even if hospitalization is good, AA might be more efficient – after all, we can't afford to hospitalize *everyone*. It's a bad objection because the authors of the study did the math and found out that hospitalization was so much better than AA that it decreased the level of further medical treatment needed and saved the health system more money than it cost.

The first good objection: like the Brandsma study, this study uses people under coercion – in this case, workers who would lose their job if they refused. Fine.

The second good objection, and this one is really interesting: *a lot of inpatient hospital rehab is AA*. That is, when you go to

an hospital for inpatient drug treatment, you attend AA groups every day, and when you leave, they make you keep going to the AA groups. In fact, the study says that "at the 12 month and 24 month assessments, the rates of AA affiliation and attendance in the past 6 months did not differ significantly among the groups." Given that the hospital patients got hospital AA + regular AA, they were actually getting *more* AA than the AA group!

So all that this study proves is that AA + more AA + other things is better than AA. There was no "no AA" group, which makes it impossible to discuss how well AA does or doesn't work. Frick.

Timko (2006) is the only study I can hesitantly half-endorse. This one has a sort of clever methodological trick to get around the limitation that doctors can't ethically refuse to refer alcoholics to treatment. In this study, researchers at a Veterans' Affairs hospital randomly assigned alcoholic patients to "referral" or "intensive referral". In "referral", the staff asked the patients to go to AA. In "intensive referral", the researchers asked REALLY NICELY for the patients to go to AA, and gave them nice glossy brochures on how great AA was, and wouldn't shut up about it, and arranged for them to meet people at their first AA meeting so they could have friends in AA, et cetera, et cetera. The hope was that more people in the "intensive referral" group would end out in AA, and ~~that indeed happened~~ scratch that, I just re-read the study and the same number of people in both groups went to AA and the intensive group actually completed a lower number of the 12 Steps on average, have I mentioned I hate all research and this entire field is terrible? But the intensive referral people were more likely to have "had a spiritual awakening" and "have a sponsor", so it was decided the study wasn't a

complete loss and when it was found the intensive referral condition had slightly less alcohol use the authors decided to declare victory.

So, whereas before we found that AA + More AA was better than AA, and that proved AA didn't work, in this study we find that AA + More AA was better than AA, and that proves AA *does* work. You know, did I say I hesitantly half-endorsed this study? Scratch that. I hate this study too.

IV.

All right, @#%^ this $@!&*. We need a *real* study, everything all lined up in a row, none of this garbage. Let's just hire half the substance abuse scientists in the country, throw a gigantic wad of money at them, give them as many patients as they need, let them take as long as they want, but barricade the doors of their office and not let them out until they've proven something important beyond a shadow of a doubt.

This was about how the scientific community felt in 1989, when they launched Project MATCH. This eight-year, $30 million dollar, multi-thousand patient trial was supposed to solve everything.

The people going into Project MATCH might have been a little overconfident. Maybe "not even *Zeus* could prevent this study from determining the optimal treatment for alcohol addiction" overconfident. This might have been a mistake.

The study was designed with three arms, one for each of the popular alcoholism treatments of the day. The first arm would be "twelve step facilitation", a form of therapy based off of Alcoholics Anonymous. The second arm would be cognitive behavioral therapy, the most bog-standard psychotherapy in the world and one which by ancient tradition must be included in any kind of study like this. The third arm would be

motivational enhancement therapy, which is a very short intervention where your doctor tells you all the reasons you should quit alcohol and tries to get you to convince yourself.

There wasn't a "no treatment" arm. This is where the overconfidence might have come in. Everyone knew alcohol treatment *worked*. Surely you couldn't dispute *that*. They just wanted to see which treatment worked best for which people. So you would enroll a bunch of different people – rich, poor, black, white, married, single, chronic alcoholic, new alcoholic, highly motivated, unmotivated – and see which of these people did best in which therapy. The result would be an algorithm for deciding where to send each of your patients. Rich black single chronic unmotivated alcoholic? We've found with $p < 0.00001$ that the best place for someone like that is in motivational enhancement therapy. Such was the dream.

So, eight years and thirty million dollars and the careers of several prestigious researchers later, the results come in, and - yeah, everyone does exactly the same on every kind of therapy (with one minor, possibly coincidental exception). Awkward.

"Everybody has won and all must have prizes!". If you're an optimist, you can say all treatments work and everyone can keep doing whatever they like best. If you're a pessimist, you might start wondering whether anything works at all.

By my understanding this is also the confusing conclusion of Ferri, Amato & Davoli (2006), the Cochrane Collaboration's attempt to get in on the AA action. Like all Cochrane Collaboration studies since the beginning of time, they find there is insufficient evidence to demonstrate the effectiveness of the intervention being investigated. This has been oft-quoted in the anti-AA literature. But by my reading, they had

no control groups and were comparing AA to different types of treatment:

> Three studies compared AA combined with other interventions against other treatments and found few differences in the amount of drinks and percentage of drinking days. Severity of addiction and drinking consequence did not seem to be differentially influenced by TSF versus comparison treatment interventions, and no conclusive differences in treatment drop out rates were reported.

So the two best sources we have – Project MATCH and Cochrane – don't find any significant differences between AA and other types of therapy. Now, to be fair, the inpatient treatment mentioned in Walsh et al wasn't included, and inpatient treatment might be the gold standard here. But sticking to various forms of outpatient intervention, they all seem to be about the same.

So, the $64,000 question: do all of them work well, or do all of them work poorly?

V.

Alcoholism studies avoid control groups like they are on fire, presumably because it's unethical not to give alcoholics treatment or something. However, there is one class of studies that doesn't have that problem. These are the ones on "brief opportunistic intervention", which is much like a turbocharged even shorter version of "motivational enhancement therapy". Your doctor tells you 'HELLO HAVE YOU CONSIDERED QUITTING ALCOHOL??!!' and sees what happens.

Brief opportunistic intervention is the most trollish medical intervention ever, because here are all these brilliant

psychologists and counselors trying to unravel the deepest mysteries of the human psyche in order to convince people to stop drinking, and then someone comes along and asks “Hey, have you tried just asking them politely?”. And it works.

Not consistently. But it works for about one in eight people. And the theory is that since it only takes a minute or two of a doctor’s time, it scales a lot faster than some sort of hideously complex hospital-based program that takes thousands of dollars and dozens of hours from everyone involved. If doctors would just spend five minutes with each alcoholic patient reminding them that no, really, alcoholism is really bad, we could cut the alcoholism rate by 1/8.

(this also works for smoking, by the way. I do this with every single one of my outpatients who smoke, and most of the time they roll their eyes, because their doctor is giving them *that speech*, but every so often one of them tells me that yeah, I’m right, they know they really should quit smoking and they’ll give it another try. I have never saved anyone’s life by dramatically removing their appendix at the last possible moment, but I have gotten enough patients to promise me they’ll try quitting smoking that I think I’ve saved at least one life just by obsessively doing brief interventions every chance I get. This is probably *the* most effective life-saving thing you can do as a doctor, enough so that if you understand it you *may* be licensed to ignore 80,000 Hours’ arguments on doctor replaceability)

Anyway, for some reason, it’s okay to do these studies with control groups. And they are so fast and easy to study that everyone studies them all the time. A meta-analysis of 19 studies is unequivocal that they definitely work.

Why do these work? My guess is that they do two things. First, they hit people who honestly didn't realize they had a problem, and inform them that they do. Second, the doctor usually says they'll "follow up on how they're doing" the next appointment. This means that a respected authority figure is suddenly monitoring their drinking and will glare at them if they stay they're still alcoholic. As someone who has gone into a panic because he has a dentist's appointment in a week and he hasn't been flossing enough – and then flossed until his teeth were bloody so the dentist wouldn't be disappointed – I can sympathize with this.

But for our purposes, the brief opportunistic intervention sets a lower bound. It says "Here's a really minimal thing that seems to work. Do other things work better than this?"

The "brief treatment" is the next step up from brief intervention. It's an hour-or-so-long session (or sometimes a couple such sessions) with a doctor or counselor where they tell you some tips for staying off alcohol. I bring it up here because the brief treatment research community spends its time doing studies that show that brief treatments are just as good as much more intense treatments. This might be most comparable to the "motivational enhancement therapy" in the MATCH study.

Chapman and Huygens (1988) find that a single interview with a health professional is just as good as six weeks of inpatient treatment (I don't know about their hospital in New Zealand, but for reference six weeks of inpatient treatment in *my* hospital costs about $40,000.)

Edwards (1977) finds that in a trial comparing "conventional inpatient or outpatient treatment complete with the full panoply of services available at a leading psychiatric

institution and lasting several months" versus an hour with a doc, both groups do the same at one and two year followup.

And so on.

All of this is starting to make my head hurt, but it's a familiar sort of hurt. It's the way my head hurts when Scott Aaronson talks about complexity classes. We have all of these different categories of things, and some of them are the same as others and others are bigger than others but we're not sure exactly where all of them stand.

We have classes "no treatment", "brief opportunistic intervention", "brief treatment", "Alcoholics Anonymous", "psychotherapy", and "inpatient".

We can prove that BOI > NT, and that AA = PT. Also that BT = IP = PT. We also have that IP > AA, which unfortunately we can use to prove a contradiction, so let's throw it out for now.

So the hierarchy of classes seems to be (NT) < (BOI) ? (BT, IP, AA, PT) - in other words, no treatment is the worst, brief opportunistic intervention is better, and then *somewhere* in there we have this class of everything else that is the same.

Can we prove that BOI = BT?

We have some good evidence for this, once again from our *Handbook*. A study in Edinburgh finds that five minutes of psychiatrist advice (brief opportunistic intervention) does the same as sixty minutes of advice plus motivational interviewing (brief treatment).

So if we take all this seriously, then it looks like every psychosocial treatment (including brief opportunistic intervention) is the same, and all are better than no treatment. This is a common finding in psychiatry and psychology – for example, all common antidepressants are better than no

treatment but work about equally well; all psychotherapies are better than no treatment but work about equally well, et cetera. It's still an open question what this says about our science and our medicine.

The strongest counterexample to this is Walsh et al which finds the inpatient hospital stay works better than the AA referral, but this study looks kind of lonely compared to the evidence on the other side. And even the authors admit they were surprised by the effectiveness of the hospital there.

And let's go back to Project MATCH. There wasn't a control group. But there were the people who dropped out of the study, who said they'd go to AA or psychotherapy but never got around to it. Cutter and Fishbain (2005) take a look at what happened to these folks. They find that the dropouts did 75% as well as the people in any of the therapy groups, and that most of the effect of the therapy groups occurred in the first week (ie people dropped out after one week did about 95% as well as people who stayed in).

To me this suggests two things. First, therapy is only a little helpful over most people quitting on their own. Second, insofar as therapy is helpful, the tiniest brush with therapy is enough to make someone think "Okay, I've had some therapy, I'll be better now". Just like with the brief opportunistic interventions, five minutes of almost anything is enough.

This is a weird conclusion, but I think it's the one supported by the data.

VI.

I should include a brief word about this giant table.

Treatment Modality	Rank order	Cumulative Evidence Score	Number of studies	%+	Mean Methodological Quotient Score	Mean Severity of Treatment Population	% Excellent
Brief interventions	1	390	34	74	13.29	2.47	53
Motivational enhancement	2	189	18	72	12.83	2.72	50
GABA agonist (Acamprosate)	3	116	5	100	11.60	3.80	20
Community Reinforcement	4.5	110	7	86	14.00	3.43	71
Self-change manual (Bibliotherapy)	4.5	110	17	59	12.65	2.59	53
Opiate antagonist (Naltrexone)	6	100	6	83	11.33	3.17	0
Behavioral self-control training	7	85	31	52	12.77	2.91	52
Behavior contracting	8	64	5	80	10.40	3.60	0
Social skills training	9	57	20	55	10.9	3.80	25
Marital therapy-Behavioral	10	44	9	56	12.33	3.44	44
Aversion therapy-Nausea	11	36	6	50	10.50	3.83	17
Case managment	12	33	5	80	10.50	3.75	0
Cognitive Therapy	13	21	10	40	10.00	3.70	10
Aversion Therapy, Covert Sensitization	14.5	18	8	38	10.88	3.50	0
Aversion therapy, Apneic	14.5	18	3	67	9.67	3.33	0
Family therapy	16	15	4	50	9.25	3.25	0
Acupuncture	17	14	3	67	9.67	3.67	0
Client-centered Counsling	18	5	8	50	11.13	3.38	13
Aversion therapy, Electrical	19	-1	18	44	11.06	3.78	17
Exercise	20	-3	3	33	11.00	2.00	0
Stress Management	21	-4	3	33	10.33	2.67	0
Antidipsotropic- Disulfiram	22	-6	27	44	11.07	3.69	26
Antidepressant-SSRI	23	-16	15	53	8.60	2.67	0
Problem Solving	24	-26	4	25	12.25	3.75	50
Lithium	25	-32	7	43	11.43	3.71	29
Marital therapy- Nonbehavioral	26	-33	8	38	12.25	3.63	25
Group process psychotherapy	27	-34	3	0	8.00	2.67	0
Functional analysis	28	-36	3	0	12.00	2.67	33
Relapse prevention	29	-38	22	36	11.73	3.23	31
Self-monitoring	30	-39	6	33	12.00	3.17	50
Hypnosis	31	-41	4	0	10.25	3.75	0
Psychedelic medication	32	-44	8	25	10.13	3.63	0
Antidipsotropic-calcium carbimide	33	-52	3	0	10.00	4.00	0
Attention Placebo	34	-59	3	0	12.33	3.33	33
Serotonin agonist	35	-68	3	0	11.33	2.33	0
Treatment as usual	36	-78	15	27	9.07	3.07	13
Twelve-step facilitation	37	-82	6	17	15.00	3.67	83
Alcoholics anonymous	38	-94	7	14	10.71	3.14	29
Anxiolytic medication	39	-98	15	27	8.13	3.40	0
Milieu therapy	40	-102	14	21	10.86	3.64	29
Antidipsotropic-metronidazole	41	-103	11	9	9.73	3.73	0
Antidepressant medication (non-SSRI)	42	-104	6	0	8.67	3.17	0
Videotape self-confrontation	43	-108	8	0	10.50	3.34	13
Relaxation training	44	-152	18	17	10/56	3.06	17
Confrontational Counseling	45	-183	12	0	10.25	3.00	33
Psychotherapy	46	-207	19	16	10.89	3.26	21
Gerneral alcoholism counseling	47	-284	23	9	11.26	3.22	22
Education (tapes, lectures, or films)	48	-443	39	13	9.77	2.44	1.5

I see it everywhere. It looks very authoritative and impressive and, of course, giant. I believe the source is Miller's Handbook of Alcoholism Treatment Approaches: Effective Alternatives, 3rd Edition, the author of which is known as a very careful scholar whom I cannot help but respect.

And the table does a good thing in discussing medications like acamprosate and naltrexone, which are very important and

effective interventions but which will not otherwise be showing up in this post.

However, the therapy part of the table looks really wrong to me.

First of all, I notice acupuncture is ranked 17 out of 48, putting in a much, *much* better showing than treatments like psychotherapy, counseling, or education. Seems fishy.

Second of all, I notice that motivational enhancement (#2), cognitive therapy (#13), and twelve-step (#37) are all about as far apart as could be, but the largest and most powerful trial ever, Project MATCH, found all three to be about equal in effectiveness.

Third of all, I notice that cognitive therapy is at #13, but psychotherapy is at #46. But cognitive therapy is a kind of psychotherapy.

Fourth of all, I notice that brief interventions, motivational enhancement, confrontational counseling, psychotherapy, general alcoholism counseling, and education are all over. But a lot of these are hard to differentiate from one another.

The table seems messed up to me. Part of it is because it is about evidence base rather than effectiveness (consider that handguns have a stronger evidence base than the atomic bomb, since they have been used many more times in much better controlled conditions, but the atomic bomb is more effective) and therefore acupuncture, which is poorly studied, can rank quite high compared to things which have even one negative study.

But part of it just seems wrong. I haven't read the full book, but I blame the tendency to conflate studies showing "X does not work better than anything else" with "X does not work".

Remember, whenever there are meta-analyses that contradict single very large well-run studies, go with the single very large well-run study, especially when the meta-analysis is as weird as this one. Project MATCH is the single very large well-run study, and it says this is balderdash. I'm guessing it's trying to use some weird algorithmic methodology to automatically rate and judge each study, but that's no substitute for careful human review.

VII.

In conclusion, as best I can tell – and it is not very well, because the studies that could really prove anything robustly haven't been done – most alcoholics get better on their own. All treatments for alcoholism, including Alcoholics Anonymous, psychotherapy, and just a few minutes with a doctor explaining why she thinks you need to quit, increase this already-high chance of recovery a small but nonzero amount. Furthermore, they are equally effective after only a tiny dose: your first couple of meetings, your first therapy session. Some studies suggest that inpatient treatment with outpatient followup may be better than outpatient treatment alone, but other studies contradict this and I am not confident in the assumption.

So does Alcoholics Anonymous work? Though I cannot say anything authoritatively, my impression is: Yes, but only a tiny bit, and for many people five minutes with a doctor may work just as well as years completing the twelve steps. As such, individual alcoholics may want to consider attending if they don't have easier options; doctors might be better off just talking to their patients themselves.

If this is true – and right now I don't have much confidence that it is, it's just a direction that weak and contradictory data

are pointing – it would be really awkward for the multibazillion-dollar treatment industry.

More worrying, I am afraid of what it would do to the War On Drugs. Right now one of the rallying cries for the anti-Drug-War movement is "treatment, not prison". And although I haven't looked seriously at the data for any drug besides alcohol. I think some data there are similar. There's very good medication for drugs – for example methadone and suboxone for opiate abuse – but in terms of psychotherapy it's mostly the same stuff you get for alcohol. Rehabs, whether they work or not, seem to serve an important sort of ritual function, where if you can send a drug abuser to a rehab you at least feel like something has been done. Deny people that ritual, and it might make prison the only politically acceptable option.

In terms of things to actually treat alcoholism, I remain enamoured of the Sinclair Method, which has done crazy outrageous stuff like conduct an experiment *with an actual control group*. But I haven't investigated enough to know whether my early excitement about them looks likely to pan out or not.

I would not recommend quitting any form of alcohol treatment that works for you, or refusing to try a form of treatment your doctor recommends, based on any of this information.

The Control Group Is Out Of Control

I.

Allan Crossman calls parapsychology the control group for science.

That is, in let's say a drug testing experiment, you give some people the drug and they recover. That doesn't tell you much until you give some other people who are taking a placebo drug you *know* doesn't work – but which they themselves believe in – and see how many of *them* recover. That number tells you how many people will recover whether the drug works or not. Unless people on your real drug do significantly better than people on the placebo drug, you haven't found anything.

On the meta-level, you're studying some phenomenon and you get some positive findings. That doesn't tell you much until you take some other researchers who are studying a phenomenon you *know* doesn't exist – but which they themselves believe in – and see how many of *them* get positive findings. That number tells you how many studies will discover positive results whether the phenomenon is real or not. Unless studies of the real phenomenon do significantly better than studies of the placebo phenomenon, you haven't found anything.

Trying to set up placebo science would be a logistical nightmare. You'd have to find a phenomenon that definitely doesn't exist, somehow convince a whole community of scientists across the world that it does, and fund them to study it for a couple of decades without them figuring out the gig.

Luckily we have a natural experiment in terms of parapsychology – the study of psychic phenomena – which most reasonable people don't believe exists but which a community of practicing scientists does and publishes papers on all the time.

The results are pretty dismal. Parapsychologists are able to produce experimental evidence for psychic phenomena about as easily as normal scientists are able to produce such evidence for normal, non-psychic phenomena. This suggests the existence of a very large "placebo effect" in science – ie with enough energy focused on a subject, you can *always* produce "experimental evidence" for it that meets the usual scientific standards. As Eliezer Yudkowsky puts it:

> Parapsychologists are constantly protesting that they are playing by all the standard scientific rules, and yet their results are being ignored – that they are unfairly being held to higher standards than everyone else. I'm willing to believe that. It just means that the standard statistical methods of science are so weak and flawed as to permit a field of study to sustain itself in the complete absence of any subject matter.

These sorts of thoughts have become more common lately in different fields. Psychologists admit to a crisis of replication as some of their most interesting findings turn out to be spurious. And in medicine, John Ioannides and others have been criticizing the research for a decade now and telling everyone they need to up their standards.

"Up your standards" has been a complicated demand that cashes out in a lot of technical ways. But there is broad agreement among the most intelligent voices I read (1, 2, 3, 4, 5) about a couple of promising directions we could go:

1. Demand very large sample size.

2. Demand replication, preferably exact replication, most preferably multiple exact replications.

3. Trust systematic reviews and meta-analyses rather than individual studies. Meta-analyses must prove homogeneity of the studies they analyze.

4. Use Bayesian rather than frequentist analysis, or even combine both techniques.

5. Stricter p-value criteria. It is far too easy to massage p-values to get less than 0.05. Also, make meta-analyses look for "p-hacking" by examining the distribution of p-values in the included studies.

6. Require pre-registration of trials.

7. Address publication bias by searching for unpublished trials, displaying funnel plots, and using statistics like "fail-safe N" to investigate the possibility of suppressed research.

8. Do heterogeneity analyses or at least observe and account for differences in the studies you analyze.

9. Demand randomized controlled trials. None of this "correlated even after we adjust for confounders" BS.

10. Stricter effect size criteria. It's easy to get small effect sizes in *anything*.

If we follow these ten commandments, then we avoid the problems that allowed parapsychology and probably a whole host of other problems we don't know about to sneak past the scientific gatekeepers.

Well, what now, motherfuckers?

II.

Bem, Tressoldi, Rabeyron, and Duggan (2014), full text available for download at the top bar of the link above, is parapsychology's way of saying "thanks but no thanks" to the idea of a more rigorous scientific paradigm making them quietly wither away.

You might remember Bem as the prestigious establishment psychologist who decided to try his hand at parapsychology and to his and everyone else's surprise got positive results. Everyone had a lot of criticisms, some of which were very very good, and the study failed replication several times. Case closed, right?

Earlier this month Bem came back with a meta-analysis of ninety replications from tens of thousands of participants in thirty three laboratories in fourteen countries confirming his original finding, $p < 1.2 * -10^{10}$, Bayes factor $7.4 * 10^{9}$, funnel plot beautifully symmetrical, p-hacking curve nice and right-skewed, Orwin fail-safe n of 559, et cetera, et cetera, et cetera.

By my count, Bem follows all of the commandments except [6] and [10]. He apologizes for not using pre-registration, but says it's okay because the studies were exact replications of a previous study that makes it impossible for an unsavory researcher to change the parameters halfway through and does pretty much the same thing. And he apologizes for the small effect size but points out that some effect sizes are legitimately very small, this is no smaller than a lot of other commonly-accepted results, and that a high enough p-value ought to make up for a low effect size.

This is *far* better than the average meta-analysis. Bem has always been pretty careful and this is no exception.

So – once again – what now, motherfuckers?

III.

In retrospect, that list of ways to fix science above was a little optimistic.

The first nine items (large sample sizes, replications, low p-values, Bayesian statistics, meta-analysis, pre-registration, publication bias, heterogeneity) all try to solve the same problem: accidentally mistaking noise in the data for a signal.

We've placed so much emphasis on not mistaking noise for signal that when someone like Bem hands us a beautiful, perfectly clear signal on a silver platter, it briefly stuns us. "Wow, of the three hundred different terrible ways to mistake noise for signal, Bem has proven beyond a shadow of a doubt he hasn't done any of them." And we get so stunned we're likely to forget that this is only part of the battle.

Bem definitely picked up a signal. The only question is whether it's a signal of psi, or a signal of poor experimental technique.

None of these five techniques even *touch* poor experimental technique – or confounding, or whatever you want to call it. If an experiment is confounded, if it produces a strong signal even when its experimental hypothesis is true, then using a larger sample size will just make that signal even stronger.

Replicating it will just reproduce the confounded results again.

Low p-values will be easy to get if you perform the confounded experiment on a large enough scale.

Meta-analyses of confounded studies will obey the immortal law of "garbage in, garbage out".

Pre-registration only assures that your study will not get any worse than it was the first time you thought of it, which may be very bad indeed.

Searching for publication bias only means you will get *all* of the confounded studies, instead of just some of them.

Heterogeneity just tells you whether all of the studies were confounded about the same amount.

Bayesian statistics, alone among these first eight, ought to be able to help with this problem. After all, a good Bayesian should be able to say "Well, I got some impressive results, but my prior for psi is very low, so this raises my belief in psi slightly, but raises my belief that the experiments were confounded *a lot*."

Unfortunately, good Bayesians are hard to come by. People like to mock Less Wrong, saying we're amateurs getting all starry-eyed about Bayesian statistics even while real hard-headed researchers who have been experts in them for years understand both their uses and their limitations. Well, maybe that's true of some researchers. But the particular ones I see talking about Bayes *here* could do with reading the Sequences. Here's Bem:

> An opportunity to calculate an approximate answer to this question emerges from a Bayesian critique of Bem's (2011) experiments by Wagenmakers, Wetzels, Borsboom, & van der Maas (2011). Although Wagenmakers et al. did not explicitly claim psi to be impossible, they came very close by setting their prior odds at 10^20 against the psi hypothesis. The Bayes Factor for our full database is approximately 10^9 in favor of the psi hypothesis (Table 1), which implies that our meta-analysis should lower their posterior odds against the psi hypothesis to 10^11

Let me shame both participants in this debate.

Bem, you are abusing Bayes factor. If Wagenmakers uses your 10^9 Bayes factor to adjust from his prior of 10^-20 to 10^-11, then what happens the next time you come up with another database of studies supporting your hypothesis? We all know you will, because you've amply proven these results weren't due to chance, so whatever factor produced these results – whether real psi or poor experimental technique – will no doubt keep producing them for the next hundred replication attempts. When those come in, does Wagenmakers have to adjust his probability from 10^-11 to 10^-2? When you get another hundred studies, does he have to go from 10^-2 to 10^7? If so, then by conservation of expected evidence he should just update to 10^+7 right now – or really to infinity, since you can keep coming up with more studies till the cows come home. But in fact he shouldn't do that, because at some point his thought process becomes "Okay, I already know that studies of this quality can consistently produce positive findings, so either psi is real or studies of this quality aren't good enough to disprove it". This point should probably happen well before he increases his probability by a factor of 10^9. See Confidence Levels Inside And Outside An Argument for this argument made in greater detail.

Wagenmakers, you are overconfident. Suppose God came down from Heaven and said in a booming voice "EVERY SINGLE STUDY IN THIS META-ANALYSIS WAS CONDUCTED PERFECTLY WITHOUT FLAWS OR BIAS, AS WAS THE META-ANALYSIS ITSELF." You would see a p-value of less than 1.2 * 10^-10 and think "I bet that was just coincidence"? And then they could do another study of the same size, also God-certified, returning exactly the same results, and you would say "I bet that was just coincidence

too"? YOU ARE NOT THAT CERTAIN OF ANYTHING. Seriously, *read the @#!$ing Sequences*.

Bayesian statistics, at least the way they are done here, aren't gong to be of much use to anybody.

That leaves randomized controlled trials and effect sizes.

Randomized controlled trials are great. They eliminate most possible confounders in one fell swoop, and are excellent at keeping experimenters honest. Unfortunately, most of the studies in the Bem meta-analysis were already randomized controlled trials.

High effect sizes are really the only thing the Bem study lacks. And it is very hard to experimental technique so bad that it consistently produces a result with a high effect size.

But as Bem points out, demanding high effect size limits our ability to detect real but low-effect phenomena. Just to give an example, many physics experiments – like the ones that detected the Higgs boson or neutrinos – rely on detecting extremely small perturbations in the natural order, over millions of different trials. Less esoterically, Bem mentions the example of aspirin decreasing heart attack risk, which it definitely does and which is very important, but which has an effect size lower than that of his psi results. If humans have some kind of *very weak* psionic faculty that under regular conditions operates poorly and inconsistently, but does indeed exist, then excluding it by definition from the realm of things science can discover would be a bad idea.

All of these techniques are about reducing the chance of confusing noise for signal. But when we think of them as the be-all and end-all of scientific legitimacy, we end up in awkward situations where they come out super-confident in a study's accuracy simply because the issue was one they

weren't geared up to detect. Because a lot of the time the problem is something more than just noise.

IV.

Wiseman & Schlitz's Experimenter Effects And The Remote Detection Of Staring is my favorite parapsychology paper ever and sends me into fits of nervous laughter every time I read it.

The backstory: there is a classic parapsychological experiment where a subject is placed in a room alone, hooked up to a video link. At random times, an experimenter stares at them menacingly through the video link. The hypothesis is that this causes their galvanic skin response (a physiological measure of subconscious anxiety) to increase, even though there is no non-psychic way the subject could know whether the experimenter was staring or not.

Schiltz is a psi believer whose staring experiments had consistently supported the presence of a psychic phenomenon. Wiseman, in accordance with nominative determinism is a psi skeptic whose staring experiments keep showing nothing and disproving psi. Since they were apparently the only two people in all of parapsychology with a smidgen of curiosity or rationalist virtue, they decided to team up and figure out why they kept getting such different results.

The idea was to plan an experiment together, with both of them agreeing on every single tiny detail. They would then go to a laboratory and set it up, again both keeping close eyes on one another. Finally, they would conduct the experiment in a series of different batches. Half the batches (randomly assigned) would be conducted by Dr. Schlitz, the other half by Dr. Wiseman. Because the two authors had very carefully standardized the setting, apparatus and procedure beforehand, "conducted by" pretty much just meant greeting the

participants, giving the experimental instructions, and doing the staring.

The results? Schlitz's trials found strong evidence of psychic powers, Wiseman's trials found no evidence whatsoever.

Take a second to reflect on how this *makes no sense*. Two experimenters in the same laboratory, using the same apparatus, having no contact with the subjects except to introduce themselves and flip a few switches – and whether one or the other was there that day completely altered the result. For a good time, watch the gymnastics they have to do to in the paper to make this sound sufficiently sensical to even get published. This is the only journal article I've ever read where, in the part of the Discussion section where you're supposed to propose possible reasons for your findings, both authors suggest maybe their co-author hacked into the computer and altered the results.

While it's nice to see people exploring Bem's findings further, *this* is the experiment people should be replicating ninety times. I expect *something* would turn up.

As it is, Kennedy and Taddonio list ten similar studies with similar results. One cannot help wondering about publication bias (if the skeptic and the believer got similar results, who cares?). But the phenomenon is sufficiently well known in parapsychology that it has led to its own host of theories about how skeptics emit negative auras, or the enthusiasm of a proponent is a necessary kindling for psychic powers.

Other fields don't have this excuse. In psychotherapy, for example, practically the only consistent finding is that whatever kind of psychotherapy the person running the study likes is most effective. Thirty different meta-analyses on the

subject have confirmed this with strong effect size (d = 0.54) and good significance (p = .001).

Then there's Munder (2013), which is a meta-meta-analysis on whether meta-analyses of confounding by researcher allegiance effect were themselves meta-confounded by meta-researcher allegiance effect. He found that indeed, meta-researchers who believed in researcher allegiance effect were more likely to turn up positive results in their studies of researcher allegiance effect ($p < .002$). It gets worse. There's a famous story about an experiment where a scientist told teachers that his advanced psychometric methods had predicted a couple of kids in their class were about to become geniuses (the students were actually chosen at random). He followed the students for the year and found that their intelligence actually increased. This was supposed to be a Cautionary Tale About How Teachers' Preconceptions Can Affect Children.

Less famous is that the same guy did the same thing with rats. He sent one laboratory a box of rats saying they were specially bred to be ultra-intelligent, and another lab a box of (identical) rats saying they were specially bred to be slow and dumb. Then he had them do standard rat learning tasks, and sure enough the first lab found very impressive results, the second lab very disappointing ones.

This scientist – let's give his name, Robert Rosenthal – then investigated three hundred forty five different studies for evidence of the same phenomenon. He found effect sizes of anywhere from 0.15 to 1.7, depending on the type of experiment involved. Note that this could also be phrased as "between twice as strong and twenty times as strong as Bem's psi effect". Mysteriously, animal learning experiments

displayed the highest effect size, supporting the folk belief that animals are hypersensitive to subtle emotional cues.

Okay, fine. Subtle emotional cues. That's way more scientific than saying "negative auras". But the question remains – what went wrong for Schlitz and Wiseman? Even if Schlitz had done everything short of saying "The hypothesis of this experiment is for your skin response to increase when you are being stared at, please increase your skin response at that time," and subjects had tried to comply, the whole point was that they didn't *know* when they were being stared at, because to find that out you'd have to be psychic. And how are these rats figuring out what the experimenters' subtle emotional cues mean anyway? *I* can't figure out people's subtle emotional cues half the time!

I know that standard practice here is to tell the story of Clever Hans and then say That Is Why We Do Double-Blind Studies. But first of all, I'm pretty sure no one does double-blind studies with rats. Second of all, I think most social psych studies aren't double blind – I just checked the first one I thought of, Aronson and Steele on stereotype threat, and it certainly wasn't. Third of all, this effect seems to be just as common in cases where it's hard to imagine how the researchers' subtle emotional cues could make a difference. Like Schlitz and Wiseman. Or like the psychotherapy experiments, where most of the subjects were doing therapy with individual psychologists and never even saw whatever prestigious professor was running the study behind the scenes.

I think it's a combination of subconscious emotional cues, subconscious statistical trickery, perfectly conscious fraud which for all we know happens much more often than detected, and things we haven't discovered yet which are at least as weird as subconscious emotional cues. But rather than

speculate, I prefer to take it as a brute fact. Studies are going to be confounded by the allegiance of the researcher. When researchers who don't believe something discover it, that's when it's worth looking into.

V.

So what exactly happened to Bem?

Although Bem looked hard to find unpublished material, I don't know if he succeeded. Unpublished material, in this context, has to mean "material published enough for Bem to find it", which in this case was mostly things presented at conferences. What about results so boring that they were never even mentioned?

And I predict people who believe in parapsychology are more likely to conduct parapsychology experiments than skeptics. Suppose this is true. And further suppose that for some reason, experimenter effect is real and powerful. That means most of the experiments conducted will support Bem's result. But this is still a weird form of "publication bias" insofar as it ignores the contrary results of hypotheticaly experiments that were never conducted.

And worst of all, maybe Bem really did do an excellent job of finding every little two-bit experiment that no journal would take. How much can we trust these non-peer-reviewed procedures?

I looked through his list of ninety studies for all the ones that were both exact replications and had been peer-reviewed (with one caveat to be mentioned later). I found only seven:

Batthyany, Kranz, and Erber: .268
Ritchie 1: 0.015
Ritchie 2: -0.219

Richie 3: -0.040
Subbotsky 1: 0.279
Subbotsky 2: 0.292
Subbotsky 3: -.399

Three find large positive effects, two find approximate zero effects, and two find large negative effects. Without doing any calculatin', this seems pretty darned close to chance for me.

Okay, back to that caveat about replications. One of Bem's strongest points was how many of the studies included were exact replications of his work. This is important because if you do your own novel experiment, it leaves a lot of wiggle room to keep changing the parameters and statistics a bunch of times until you get the effect you want. This is why lots of people want experiments to be preregistered with specific committments about what you're going to test and how you're going to do it. These experiments weren't preregistered, but conforming to a previously done experiment is a pretty good alternative.

Except that I think the criteria for "replication" here were exceptionally loose. For example, Savva et al was listed as an "exact replication" of Bem, but it was performed in 2004 – seven years before Bem's original study took place. I know Bem believes in precognition, but that's going *too far*. As far as I can tell "exact replication" here means "kinda similar psionic-y thing". Also, Bem classily lists his own experiments as exact replications of themselves, which gives a big boost to the "exact replications return the same results as Bem's original studies" line. I would want to see much stricter criteria for replication before I relax the "preregister your trials" requirement.

(Richard Wiseman – the same guy who provided the negative aura for the Wiseman and Schiltz experiment – has started a pre-register site for Bem replications. He says he has received five of them. This is very promising. There is also a separate pre-register for parapsychology trials in general. I am both extremely pleased at this victory for good science, and ashamed that my own field is apparently behind parapsychology in the "scientific rigor" department)

That is my best guess at what happened here – a bunch of poor-quality, peer-unreviewed studies that weren't as exact replications as we would like to believe, all subject to mysterious experimenter effects.

This is not a criticism of Bem or a criticism of parapsychology. It's something that is inherent to the practice of meta-analysis, and even more, inherent to the practice of science. Other than a few very exceptional large medical trials, there is not a study in the world that would survive the level of criticism I am throwing at Bem right now.

I think Bem is wrong. The level of criticism it would take to prove a wrong study wrong is higher than that almost any existing study can withstand. That is not encouraging for existing studies.

VI.

The motto of the Royal Society – Hooke, Boyle, Newton, some of the people who arguably invented modern science – was *nullus in verba*, "take no one's word".

This was a proper battle cry for seventeenth century scientists. Think about the (admittedly kind of mythologized) history of Science. The scholastics saying that matter was this, or that, and justifying themselves by long treatises about how based on A, B, C, the word of the Bible, Aristotle, self-evident first

principles, and the Great Chain of Being all clearly proved their point. Then other scholastics would write different long treatises on how D, E, and F, Plato, St. Augustine, and the proper ordering of angels all indicated that clearly matter was something different. Both groups were pretty sure that the other had make a subtle error of reasoning somewhere, and both groups were perfectly happy to spend centuries debating exactly which one of them it was.

And then Galileo said "Wait a second, instead of debating exactly how objects fall, let's just drop objects off of something really tall and see what happens", and after that, Science.

Yes, it's kind of mythologized. But like all myths, it contains a core of truth. People are terrible. If you let people debate things, they will do it forever, come up with horrible ideas, get them entrenched, play politics with them, and finally reach the point where they're coming up with theories why people who disagree with them are probably secretly in the pay of the Devil.

Imagine having to conduct the global warming debate, except that you couldn't appeal to scientific consensus and statistics because scientific consensus and statistics hadn't been invented yet. In a world without science, *everything* would be like that.

Heck, just look at *philosophy*.

This is the principle behind the Pyramid of Scientific Evidence. The lowest level is your personal opinions, no matter how ironclad you think the logic behind them is. Just above that is expert opinion, because no matter how expert someone is they're still only human. Above that is anecdotal evidence and case studies, because even though you're finally

getting out of people's heads, it's still possible for the content of people's heads to influence which cases they pay attention to. At each level, we distill away more and more of the human element, until presumably at the top the dross of humanity has been purged away entirely and we end up with pure unadulterated reality.

The Pyramid of Scientific Evidence

And for a while this went *well*. People would drop things off towers, or see how quickly gases expanded, or observe chimpanzees, or whatever.

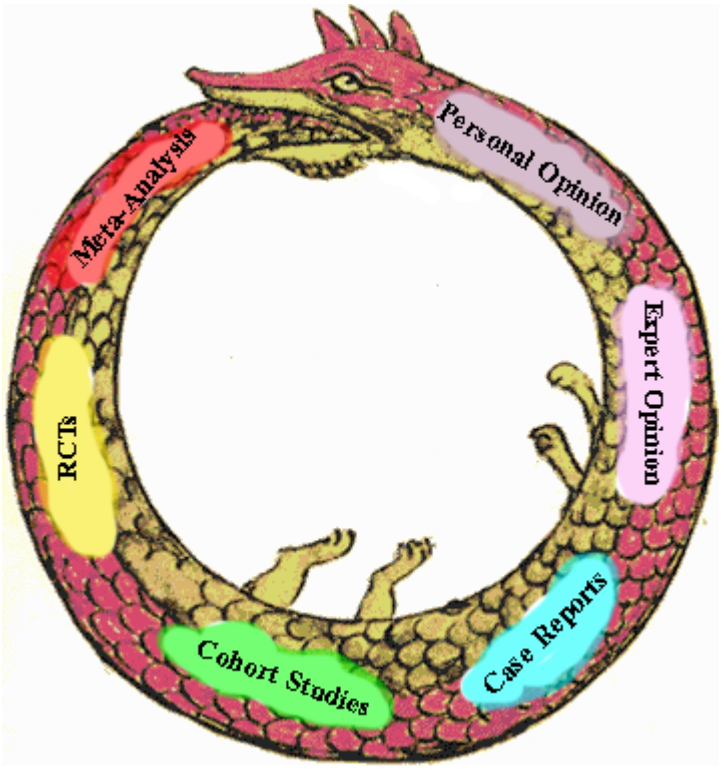
Then things started getting more complicated. People started investigating more subtle effects, or effects that shifted with the observer. The scientific community became bigger, everyone didn't know everyone anymore, you needed more journals to find out what other people had done. Statistics became more complicated, allowing the study of noisier data but also bringing more peril. And a lot of science done by smart and honest people ended up being wrong, and we needed to figure out exactly which science that was.

And the result is a lot of essays like this one, where people who think they're smart take one side of a scientific "controversy" and say which studies you should believe. And then other people take the other side and tell you why you

should believe different studies than the first person thought you should believe. And there is much argument and many insults and citing of authorities and interminable debate for, if not centuries, at least a pretty long time.

The highest level of the Pyramid of Scientific Evidence is meta-analysis. But a lot of meta-analyses are crap. This meta-analysis got p < 1.2 * 10^-10 for a conclusion I'm pretty sure is false, and *it isn't even one of the crap ones*. Crap meta-analyses look more like this, or even worse.

How do I know it's crap? Well, I use my personal judgment. How do I know my personal judgment is right? Well, a smart well-credentialed person like James Coyne agrees with me. How do I know James Coyne is smart? I can think of lots of cases where he's been right before. How do I know those count? Well, John Ioannides has published a lot of studies analyzing the problems with science, and confirmed that cases like the ones Coyne talks about are pretty common. Why can I believe Ioannides' studies? Well, there have been good meta-analyses of them. But how do I know if those meta-analyses are crap or not? Well…



The Ouroboros of Scientific Evidence

Science! YOU WERE THE CHOSEN ONE! It was said that you would destroy reliance on biased experts, not join them! Bring balance to epistemology, not leave it in darkness!

I LOVED YOU!!!!

Edit: Conspiracy theory by Andrew Gelman

The Cowpox of Doubt

I remember hearing someone I know try to explain rationality to his friends.

He started with "It's important to have correct beliefs. You might think this is obvious, but think about creationists and homeopaths and people who think the moon landing was a hoax." And then further on in this vein.

And I thought: "NO NO NO NO NO NO NO!"

I will make a confession. Every time someone talks about the stupidity of creationists, moon-hoaxers, and homeopaths, I cringe.

It's not that moon-hoaxers, homeopaths et al aren't dumb. They are. It's not even that these people don't do real harm. They do.

(although probably less than people think; people rarely stop conventional treatment in favor of homeopathy, and both a popular website and a review article have a really hard time finding more than a handful of people genuinely harmed by it. Moon hoaxes seem even less dangerous, unless of course you are standing near Buzz Aldrin when you talk about them.)

What annoys me about the people who harp on moon-hoaxing and homeopathy – without any interest in the rest of medicine or space history – is that it seems like an attempt to Other irrationality.

(yes, I did just use "other" as a verb. Maybe I've been hanging around Continental types too much lately.)

It's saying "Look, over here! It's irrational people, believing things that we can instantly dismiss as dumb. Things we feel

no temptation, not one bit, to believe. It must be that they are defective and we are rational."

But to me, the rationality movement is about Self-ing irrationality.

(yes, I did just use "self" as a verb. I don't even have the excuse of it being part of a philosophical tradition)

It is about realizing that you, yes you, might be wrong about the things that you're most certain of, and nothing can save you except maybe extreme epistemic paranoia.

Talking about moon-hoaxers and homeopaths too much, at least the way we do it, is *counterproductive* to this goal. Throw examples of obviously stupid false beliefs at someone, and they start thinking all false beliefs are obvious. Give too many examples of false beliefs that aren't tempting to them, and they start believing they're immune to temptation.

And it raises sloppiness to a virtue.

Take homeopathy. I can't even count the number of times I've heard people say: "Homeopaths don't realize beliefs require evidence. No study anywhere has ever found homeopathy to be effective!"

But of course dozens of studies have found homeopathy to be effective.

"Well, sure, but they weren't double-blind! What you don't realize is that there can be placebo effects from…"

But of course many of these studies have been large double-blinded randomized controlled trials, or even meta-analyses of such.

"Okay, but not published in reputable journals."

Is *The Lancet* reputable enough for you?

“But homeopaths don’t even realize that many of their concoctions don’t contain even a single molecule of active substance!”

But of course almost all homeopaths realize this and their proposed mechanism for homeopathic effects not only survives this criticism but relies upon it.

“But all doctors and biologists agree that homeopathy doesn’t work!”

Have you ever spent the five seconds it would take to look up a survey of what percent of doctors and biologists believe homeopathy doesn’t work? Or are you just assuming that’s true because someone on your side told you so and it seems right?

I am of course being mean here. Being open-minded to homeopaths – reading all the research carefully, seeking out their own writings so you don’t accidentally straw-man them, double-checking all of your seemingly “obvious” assumptions – would be a waste of your time.

And someone who demands that you be open-minded about homeopathy would not be your friend. They would probably be a shill for homeopathy and best ignored.

But this is exactly the problem!

The more we concentrate on homeopathy, and moon hoaxes, and creationism – the more people who have never felt any temptation towards these beliefs go through the motions of “debunk”-ing them a hundred times to one another for fun – the more we are driving home the message that these are a representative sample of the kinds of problems we face.

And the more we do that, the more we are training people to make the correct approach to homeopathy – ignoring poor

research and straw men on your own side while being very suspicious of anyone who tells us to be careful – their standard approach to any controversy.

And then we get people believing all sorts of shoddy research – because after all, the world is divided between things like homeopathy that Have Never Been Supported By Any Evidence Ever, and things like conventional medicine that Have Studies In Real Journals And Are Pushed By Real Scientists.

Or losing all subtlety and moderation in their political beliefs, never questioning their own side's claims, because the world is divided between People Like Me Who Know The Right Answer, and Shills For The Other Side Who Tell Me To Be Open-Minded As Part Of A Trap.

This post was partly inspired by Gruntled and Hinged's You Probably Don't Want Peer-Reviewed Evidence For God (actually, I started writing it before that was published – but since Bem has published evidence showing psi exists, I must have just been precognitively inspired by it). But there's another G&H post that retrocausally got me thinking even more.

Inoculation is when you use a weak pathogen like cowpox to build immunity against a stronger pathogen like smallpox. The inoculation effect in psychology is when a person, upon being presented with several weak arguments against a proposition, becomes immune to stronger arguments against the same position.

Tell a religious person that Christianity is false because Jesus is just a blatant ripoff of the warrior-god Mithras and they'll open up a Near Eastern history book, notice that's not true at all, and then be that much more skeptical of the next argument

against their faith. “Oh, atheists. Those are those people who think stupid things like Jesus = Mithras. I already figured out they’re not worth taking seriously.” Except on a deeper level that precedes and is immune to conscious thought.

So we take the intelligent Internet-reading public, and we throw a bunch of incredibly dumb theories at them – moon-hoaxism, homeopathy, creationism, anti-vaxxing, lizard people, that one guy who thought the rapture would come a couple years ago, whatever. And they are easily debunked, and the stuff you and all your friends believed was obviously true is, in fact, obviously true, and any time you spent investigating whether you were wrong is time you wasted.

And I worry that we are vaccinating people against reading the research for themselves instead of trusting smarmy bloggers who talk about how stupid the other side is.

That we are vaccinating people against thinking there might be important truths on both sides of an issue.

That we are vaccinating people against understanding how “scientific evidence” is a really complicated concept, and that many things that are in peer-reviewed journals will later turn out to be wrong.

That we are vaccinating people against the idea that many theories they find absurd or repugnant at first will later turn out to be true, because nature doesn’t respect our feelings.

That we are vaccinating people against *doubt*.

And maybe this is partly good. It’s probably a good idea to trust your doctor and also a good idea to trust your climatologist, and rare is the field where I would feel comfortable challenging expert consensus completely.

But there's also this problem of hundreds of different religions and political ideologies, and most people are born into ones that are at least somewhat wrong. That makes this capacity for real doubt – doubting something even though all your family and friends is telling you it's obviously true and you must be an idiot to question it at all – a tremendously important skill. It's especially important for the couple of rare individuals who will be in a position to cause a paradigm shift in a science by doubting one of its fundamental assumptions.

I don't think that reading about lizard people or creationism will affect people's ability to distinguish between, let's say, cyclic universe theory versus multiverse theory, or other equally dispassionate debates.

But if ever you ever need to have a true crisis of faith, then any time you spend thinking about homeopathy and moon hoaxes beyond the negligible effect they have on your life will be time spent learning exactly the wrong mental habits.

The Skeptic's Trilemma

Followup to: Talking Snakes: A Cautionary Tale

Related to: Explain, Worship, Ignore

Skepticism is like sex and pizza: when it's good, it's very very good, and when it's bad, it's still pretty good.

It really is hard to dislike skeptics. Whether or not their rational justifications are perfect, they are doing society a service by raising the social cost of holding false beliefs. But there is a failure mode for skepticism. It's the same as the failure mode for so many other things: it becomes a blue vs. green style tribe, demands support of all 'friendly' arguments, enters an affective death spiral, and collapses into a cult.

What does it look like when skepticism becomes a cult? Skeptics become more interested in supporting their "team" and insulting the "enemy" than in finding the truth or convincing others. They begin to think "If a assigning .001% probability to Atlantis and not accepting its existence without extraordinarily compelling evidence is good, then assigning 0% probability to Atlantis and refusing to even consider any evidence for its existence must be *great*!" They begin to deny any evidence that seems pro-Atlantis, and cast aspersions on the character of anyone who produces it. They become anti-Atlantis fanatics.

Wait a second. There is no lost continent of Atlantis. How do I know what a skeptic would do when confronted with evidence for it? For that matter, why do I care?

Way back in 2007, Eliezer described the rationalist equivalent of Abort, Retry, Fail: the trilemma of Explain, Worship, Ignore. Don't understand where rain comes from? You can try

to explain it as part of the water cycle, although it might take a while. You can worship it as the sacred mystery of the rain god. Or you can ignore it and go on with your everyday life.

So someone tells you that Plato, normally a pretty smart guy, wrote a long account of a lost continent called Atlantis complete with a bunch of really specific geographic details that seem a bit excessive for a meaningless allegory. Plato claims to have gotten most of the details from a guy called Solon, legendary for his honesty, who got them from the Egyptians, who are known for their obsessive record-keeping. This seems interesting. But there's no evidence for a lost continent anywhere near the Atlantic Ocean, and geology tells us continents can't just go missing.

One option is to hit Worship. Between the Theosophists, Edgar Cayce, the Nazis, and a bunch of well-intentioned but crazy amateurs including a U.S. Congressman, we get a supercontinent with technology far beyond our wildest dreams, littered with glowing crystal pyramids and powered by the peaceful and eco-friendly mystical wisdom of the ancients, source of all modern civilization and destined to rise again to herald the dawning of the Age of Aquarius.

Or you could hit Ignore. I accuse the less pleasnt variety of skeptic of taking this option. Atlantis is stupid. Anyone who believes it is stupid. Plato was a dirty rotten liar. Any scientist who finds anomalous historical evidence suggesting a missing piece to the early history of the Mediterranean region is also a dirty rotten liar, motivated by crazy New Age beliefs, and should be fired. Anyone who talks about Atlantis is the Enemy, and anyone who denies Atlantis gains immediate access to our in-group and official Good Rational Scientific Person status.

Spyridon Marinatos, a Greek archaeologist who really deserves more fame than he received, was a man who hit Explain. The geography of Plato's Atlantis, a series of concentric circles of land and sea, had been derided as fanciful; Marinatos noted[1] that it matched the geography of the Mediterranean island of Santorini quite closely. He also noted that Santorini had a big volcano right in the middle and seemed somehow linked to the Minoan civilization, a glorious race of seafarers who had mysteriously collapsed a thousand years before Plato. So he decided to go digging in Santorini. And he found…

…the lost city of Atlantis. Well, I'm making an assumption here. But the city he found was over four thousand years old, had a population of over ten thousand people at its peak, boasted three-story buildings and astounding works of art, and had hot and cold running water - an unheard-of convenience that it shared with the city in Plato's story. For the Early Bronze Age, that's *darn*ed impressive. And like Plato's Atlantis, it was destroyed in a single day. The volcano that loomed only a few miles from its center went off around 1600 BC, utterly burying it and destroying its associated civilization. No one knows what happened to the survivors, but the most popular theory is that some fled to Egypt[2], with which the city had flourishing trade routes at its peak.

The Atlantis = Santorini equivalence is still controversial, and the point of this post isn't to advocate for it. But just look at the difference between Joe Q. Skeptic and Dr. Marinatos. Both were rightly skeptical of the crystal pyramid story erected by the Atlantis-worshippers. But Joe Q. Skeptic considered the whole issue a nuisance, or at best a way of proving his intellectual superiority over the believers. Dr. Marinatos saw

an honest mystery, developed a theory that made testable predictions, then went out and started digging.

The fanatical skeptic, when confronted with some evidence for a seemingly paranormal claim, says “Wow, that’s stupid.” It’s a soldier on the opposing side, and the only thing to be done with it is kill it as quickly as possible. The wise skeptic, when confronted with the same evidence, says “Hmmm, that’s interesting.”

Did people at Roswell discovered the debris of a strange craft made of seemingly otherworldly material lying in a field, only to be silenced by the government later? You can worship the mighty aliens who are cosmic bringers of peace. You can ignore it, because UFOs don’t exist so the people are clearly lying. Or you can search for an explanation until you find that the government was conducting tests of Project Mogul in that very spot.

Do thousands of people claim that therapies with no scientific basis are working? You can worship alternative medicine as a natural and holistic alternative to stupid evil materialism. You can ignore all the evidence for their effectiveness. Or you can shut up and discover the placebo effect, explaining the lot of them in one fell swoop.

Does someone claim to see tiny people, perhaps elves, running around and doing elvish things? You can call them lares and worship them as household deities. You can ignore the person because he’s an obvious crank. Or you can go to a neurologist, and he’ll explain that the person’s probably suffering from Charles Bonnet Syndrome.

All unexplained phenomena are real. That is, they’re real unexplained phenomena. The explanation may be prosaic, like that people are gullible. Or it may be an entire four thousand

year old lost city of astounding sophistication. But even "people are gullible" can be an interesting explanation if you're smart enough to make it one. There's a big difference between "people are gullible, so they believe in stupid things like religion, let's move on" and a complete list of the cognitive biases that make explanations involving agency and intention more attractive than naturalistic explanations to a naive human mind. A sufficiently intelligent thinker could probably reason from the mere existence of religion all the way back to the fundamentals of evolutionary psychology.

This I consider a specific application of a more general rationalist technique: not prematurely dismissing things that go against your worldview. There's a big difference between dismissing that whole Lost Continent of Atlantis story, and *prematurely* dismissing it. It's the difference between discovering an ancient city and resting smugly satisfied that you don't have to.

Footnotes

1: I may be unintentionally sexing up the story here. I read a book on Dr. Marinatos a few years ago, and I know he did make the Santorini-Atlantis connection, but I don't remember whether he made it before starting his excavation, or whether it only clicked during the dig (and the Internet is silent on the matter). If it was the latter, all of my moralizing about how wonderful it was that he made a testable prediction falls a bit flat. I should have used another example where I knew for sure, but this story was too perfect. Mea culpa.

2: I don't include it in the main article because it is highly controversial and you have to fudge some dates for it to really work out, but here is a Special Bonus Scientific Explanation of

a Paranormal Claim: the eruption of this same supervolcano in 1600 BC caused the series of geologic and climatological catastrophes recorded in the Bible as the Ten Plagues of Egypt. However, I specify that I'm including this because it's fun to think about rather than because there's an especially large amount of evidence for it.

If You Can't Make Predictions, You're Still in a Crisis

A *New York Times* article by Northeastern University professor Lisa Feldman Barrett claims that Psychology Is Not In Crisis:

> Is psychology in the midst of a research crisis?
>
> An initiative called the Reproducibility Project at the University of Virginia recently reran 100 psychology experiments and found that over 60 percent of them failed to replicate — that is, their findings did not hold up the second time around. The results, published last week in Science, have generated alarm (and in some cases, confirmed suspicions) that the field of psychology is in poor shape.
>
> But the failure to replicate is not a cause for alarm; in fact, it is a normal part of how science works.
>
> Suppose you have two well-designed, carefully run studies, A and B, that investigate the same phenomenon. They perform what appear to be identical experiments, and yet they reach opposite conclusions. Study A produces the predicted phenomenon, whereas Study B does not. We have a failure to replicate.
>
> Does this mean that the phenomenon in question is necessarily illusory? Absolutely not. If the studies were well designed and executed, it is more likely that the phenomenon from Study A is true only under certain conditions. The scientist's job now is to figure out what those conditions are, in order to form new and better hypotheses to test […]

> When physicists discovered that subatomic particles didn't obey Newton's laws of motion, they didn't cry out that Newton's laws had "failed to replicate." Instead, they realized that Newton's laws were valid only in certain contexts, rather than being universal, and thus the science of quantum mechanics was born […]
>
> Science is not a body of facts that emerge, like an orderly string of light bulbs, to illuminate a linear path to universal truth. Rather, science (to paraphrase Henry Gee, an editor at Nature) is a method to quantify doubt about a hypothesis, and to find the contexts in which a phenomenon is likely. Failure to replicate is not a bug; it is a feature. It is what leads us along the path — the wonderfully twisty path — of scientific discovery.

Needless to say, I disagree with this rosy assessment.

The first concern is that it ignores publication bias. One out of every twenty studies will be positive by pure chance – more if you're willing to play fast and loose with your methods. Probably quite a lot of the research we see is that 1/20. Then when it gets replicated in a preregistered trial, it fails. This is not because the two studies were applying the same principle to different domains. It's because the first study posited something that simply wasn't true, in any domain. This may be *the outright majority* of replication failures, and you can't just sweep this under the rug with paeans to the complexity of science.

The second concern is experimenter effects. Why do experimenters who believe in and support a phenomenon usually find it occurs, and experimenters who doubt the phenomenon usually find that it doesn't? That's easy to explain through publication bias and other forms of bias, but if

we're just positing that there are some conditions where it does work and others where it doesn't, the ability of experimenters to so often end out in the conditions that flatter their preconceptions is a remarkable coincidence.

The third and biggest concern is the phrase "it is more likely". Read that sentence again: "If the studies were well designed and executed, it is more likely that the phenomenon from Study A is true only under certain conditions [than that it is illusory]". Really? *Why?* This is exactly the thing that John Ioannidis has spent so long arguing against! Suppose that I throw a dart at the Big Chart O' Human Metabolic Pathways and when it hits a chemical I say "This! This is the chemical that is the key to curing cancer!". Then I do a study to check. There's a 5% chance my study comes back positive by coincidence, an even higher chance that a biased experimenter can hack it into submission, but a much smaller chance that out of the thousands of chemicals I just so happened to pick the one that really does cause cancer. So if my study comes back positive, but another team's study comes back negative, it's not "more likely" that my chemical does cure cancer but only under certain circumstances. Given the base rate – that most hypotheses are false – it's more likely that I accidentally proved a false hypothesis, a very easy thing to do, and now somebody else is correcting me.

Given that many of the most famous psychology results are either extremely counterintuitive or highly politically motivated, there is *no reason at all* to choose a prior probability of correctness such that we should try to reconcile our prior belief in them with a study showing they don't work. It would be like James Randi finding Uri Geller can't bend spoons, and saying "Well, he bent spoons other times, but not around Randi, let's try to figure out what feature of Randi's

shows interferes with the magic spoon-bending rays". I am not saying that we *shouldn't* try to reconcile results and failed replications of those results, but we should do so in an informed Bayesian way instead of automatically assuming it's "more likely" that they deserve reconciliation.

Yet even ignoring the publication bias, and the low base rates, and the statistical malpractice, and the couple of cases of outright falsification, and concentrating on the ones that really *are* differences in replication conditions, this is *still* a crisis.

A while ago, Dijksterhuis and van Knippenberg published a famous priming study showing that people who spend a few minutes before an exam thinking about brilliant professors will get better grades; conversely, people who spend a few minutes thinking about moronic soccer hooligans will get worse ones. They did four related experiments, and all strongly confirmed their thesis. A few years later, Shanks et al tried to replicate the effect and couldn't. They did the same four experiments, and none of them replicated at all. What are we to make of this?

We could blame differences in the two experiments' conditions. But the second experiment made every attempt to match the conditions of the first experiment as closely as possible. Certainly they didn't do anything idiotic, like switch from an all-female sample to an all-male sample. So if we want to explain the difference in results, we have to think on the level of tiny things that the replication team wouldn't have thought about. The color of the wallpaper in the room where the experiments were taking place. The accents of the scientists involved. The barometric pressure on the day the study was conducted.

We could laboriously test the effect of wallpaper color, scientist accent, and barometric pressure on priming effects, but it would be extraordinarily difficult. Remember, we've already shown that two well-conducted studies can get diametrically opposite results. Who is to say that if we studied the effect of wallpaper color, the first study wouldn't find that it made a big difference and the second study find that it made no difference at all? What we'd probably end out with is a big conflicting morass of studies that's even more confusing than the original smaller conflicting morass.

But as far as I know, nobody is doing this. There is not enough psychology to devote time to teasing out the wallpaper-effect from the barometric-pressure effect on social priming. Especially given that maybe at the end of all of these dozens of teasing-apart studies we would learn nothing. And that quite possibly the original study was simply wrong, full stop.

Since we have not yet done this, and don't even know if it would work, we can expect even strong and well-accepted results not to apply in even very slightly different conditions. But that makes claims of scientific understanding very weak. When a study shows that Rote Memorization works better than New Math, we hope this means we've discovered something about human learning and we can change school curricula to reflect the new finding and help children learn better. But if we fully expect that the next replication attempt will show New Math is better than Rote Memorization, then that plan goes down the toilet and we shouldn't ask schools to change their curricula at all, let alone claim to have figured out deep truths about the human mind.

Barrett states that psychology is not in crisis, because it's in a position similar to physics, where gravity applies at the macroscopic level but not the microscopic level. But if you ask

a physicist to predict whether an apple will fall up or down, she will say “Down, obviously, because we’re talking about the macroscopic level.” If you ask a psychologist to predict whether priming a student with the thought of a brilliant professor will make them do better on an exam or not, the psychologist will have no idea, because she won’t know what factors cause the prime to work sometimes and fail other times, or even whether it really ever works at all. She will be at the level of a physicist who says “Apples sometimes fall down, but equally often they fall up, and we can’t predict which any given apple will do at any given time, and we don’t know why – but our field is not in crisis, because in theory some reason should exist. Maybe.”

If by physics you mean “the practice of doing physics experiments”, then perhaps that is justified. If by physics you mean “a collection of results that purport to describe physical reality”, then it’s clear you don’t actually have any.

So the *Times* article is not an argument that psychology is not in crisis. It is, at best, an IOU, saying that we should keep doing psychology because maybe if we work really hard we will reach a point where the crisis is no longer so critical.

On the other hand, there’s one part of this I agree with entirely. I don’t think we can do a full post-mortem on every failed replication. But we ought to do them on *some* failed replications. Right now, failed replications are *deeply mysterious*. Is it really things like the wallpaper color or barometric pressure? Or is it more sinister things, like failure to double-blind, or massive fraud? How come this keeps happening to us? I don’t know. If we could solve one or two of these, we might at least know what we’re up against.

IV. Medicine, Therapy, and Human Enhancement

Scientific Freud

In this month's *American Journal of Psychiatry*: The Efficacy of Cognitive-Behavioral Therapy and Psychodynamic Therapy in the Outpatient Treatment of Major Depression: A Randomized Clinical Trial. It's got more than just a catchy title. It also demonstrates that…

Wait. Before we go further, a moment of preaching.

Skepticism and metaskepticism seem to be two largely separate skills.

That is, the ability to debunk the claim "X is true" does not generalize to the ability to debunk the claim "X has been debunked".

I have this problem myself.

I was taught the following foundation myth of my field: in the beginning, psychiatry was a confused amalgam of Freud and Jung and Adler and anyone else who could afford an armchair to speculate in. People would say things like that neurosis was caused by wanting to have sex with your mother, or by secretly wanting a penis, or goodness only knows what else. Then someone had the bright idea that beliefs ought to be based on evidence! Study after study proved the psychoanalysts' bizarre castles were built on air, and the Freudians were banished to the outer darkness. Their niche was filled by newer scientific psychotherapies with a robust evidence base, such as cognitive behavioral therapy and [mumble]. And thus was the empire forged.

Now normally when I hear something this convenient, I might be tempted to make sure that there were *actual* studies this was based on. In this case, I dropped the ball. The Heroic

Foundation Myth isn't a *claim*, I must have told myself. It's a *debunking*. To be skeptical of the work of fellow debunkers would be a violation of professional courtesy!

The AJP article above is interesting because as far as I know it's the largest study ever to compare Freudian and cognitive-behavioral therapies. It examined both psychodynamic therapy (a streamlined, shorter-term version of Freudian psychoanalysis) and cognitive behavioral therapy on 341 depressed patients. It found – using a statistic called noninferiority which I don't entirely understand – that CBT was no better than psychoanalysis. In fact, although the study wasn't designed to demonstrate this, just by eyeballing it looks like psychoanalysis did nonsignificantly better. The journal's editorial does a good job putting the result in context.

This follows on the heels of several other studies and meta-analyses finding no significant difference between the two therapies, including, another in depression, yet another in depression, still another in depression, one in generalized anxiety disorder and one in general. This study by meta-analysis celebrity John Ioannidis also seems to incidentally find no difference between psychodynamics and CBT, although that wasn't quite what it was intended to study and it's probably underpowered to detect a difference.

(other analyses do show a difference, for example Tolin et al, but the studies they draw from tend to be much smaller than this latest and in any case are starting to look increasingly lonely.)

Suppose we accept the conclusion in this and many other articles that psychodynamic therapy is equivalent to cognitive-behavioral therapy. Do we have to accept that Freud was right after all?

Well, one man's modus ponens is another man's modus tollens. The other possible conclusion is that cognitive-behavioral therapy doesn't really work either.

If parapsychology is the control group for science, Freudian psychodynamics really ought to be the control group for psychotherapy. Although I know some really intelligent people who take it seriously, to me it seems so outlandish, such a shot-in-the-dark in a low-base-rate-of-success environment, that we can dismiss it out of hand and take any methodology that approves of it to be more to the shame of the methodology than to the credit of the therapy.

But what about the evidence base for cognitive behavioral therapy over placebo? Or, for that matter, the evidence base for psychoanalysis over placebo?

Part of the problem may be what exactly is used as placebo psychotherapy. In many studies, it's just getting random people to talk to patients. This makes intuitive sense as a placebo therapy, but it seems vulnerable to unblinding – people usually have some expectation of what psychotherapy is like, and undirected conversation about problems might not match it. Or if the placebo therapists are not professionals, they may be less confident in talking to people about their mental health problems, more awkward, less charismatic, or otherwise not the sort of people who would make it in the therapy profession. So now a lot of people are coalescing around the idea that all therapy studies done against these kinds of placebo therapy are fundamentally flawed.

Studies that compare what are called "bona fide psychotherapies" – two therapies both done by real therapists with real training – tend to have a lot more trouble finding differences. This has led to what is called the Dodo Bird

Verdict, after an obscure Alice in Wonderland reference I feel vaguely bad for not getting: that psychotherapies work by having a charismatic, caring person listen to your problems and then do ritualistic psychotherapy-sounding things to you, but not by any of the exercises or theories of the specific therapy itself.

Then the question becomes: if the Dodo Bird Verdict and the active placebo problem and so on are equally true of all psychotherapies and all psychotherapy studies, how come everyone become convinced that cognitive behavioral therapy passed the evidence test and psychoanalysis failed it?

And the answer is *the CBT people did studies and the psychoanalysts didn't*.

That's it. It may be, it probably is, that any study would have come back positive. But only the cognitive behavioral people bothered to perform any. And by the time the situation was rectified and the psychoanalysts had (positive) studies of their own to hold up, "everyone knew" that CBT was evidence-based and psychoanalysis wasn't.

This seems like another case of doctors not understanding that there are two different types of "no evidence".

I should qualify this sweeping condemnation. I believe a few very basic therapies that address specific symptoms in very simple ways will work. For example, exposure therapy – where you treat someone's fear of snakes by throwing snakes at them until they realize it's harmless – is extremely and undeniably effective. Some versions of CBT for anxiety and DBT for borderline also seem to just be basic coping skills about getting some distance from your emotions. I think it's likely that these have some small effects (I know a study above found no effect for CBT on anxiety, but it was by a

notorious partisan of psychoanalysis and I will temporarily defy the data).

But anything more complicated than that, anything based on an overarching theory of How The Mind Works, and I intuitively side with the Dodo Bird Verdict. And I think the evidence backs me up.

EDIT: Do *not* stop going to psychotherapy after reading this post! All psychotherapies, including placebo psychotherapies, are much better than nothing at all (kinda like how all psychiatric medications, including placebo medications, are much better than nothing at all).

Sleep – Now by Prescription

Ramelteon isn't a bad drug. It's just that its very existence stands as a condemnation of the entire medical system.

All sleep medications have to straddle a very fine line between "idiotically dangerous" and "laughably ineffective", and Ramelteon manages better than most. It outperforms placebo, it's not addictive, it won't sap your ability to sleep without it, and it doesn't screw up your brain so badly that its unofficial mascot is a hallucinatory walrus.

How does it do it? Ramelteon is the first melatonergic drug, selectively binding to MT-1 and MT-2 melatonin receptors. Binding to melatonin receptors presumably mimics the effect of the natural hormone melatonin which is believed to serve a sleep-promoting role.

Now, you might ask yourself – the natural hormone melatonin is available as an over-the-counter supplement costing a couple cents per pill in every drug store, and provably quite safe and effective. Why would anyone go through the trouble of creating a drug that mimics its action? Especially if a month's supply of the drug costs around $100 – which it does.

The answer is: *I have no idea and I'm pretty sure no one else does either*.

Wikipedia says of Ramelteon that:

> In a double-blind multicenter trial, Ramelteon did reduce the time to fall asleep by approximately 15–20 minutes, at 8 mg and 16 mg doses after four weeks compared to placebo (approx. 29-32 versus 48 minutes) Total sleep

> time improved about 40 minutes, however, this was identical to improvement with placebo at the end of trial

A meta-analysis of melatonin says:

> Our meta-analysis demonstrated melatonin had a significant benefit in reducing sleep latency. Subjects randomly assigned to melatonin fell asleep 7 minutes earlier on average than subjects receiving placebo…in the random effects model, sleep latency was reduced by over 10 minutes

Sleep latency is a tough statistic to work with, because it depends a lot on how quickly the people in your trial got to sleep in the first place. If the study population is chronic insomniacs who take an hour to fall asleep each night, a good drug might be able to reduce that by 30 minutes. If the study population is normal youth who fall asleep within ten minutes, needless to say your drug isn't going to be able to do 30 minutes better.

So, for example, it's easy to find a melatonin trial that finds a very impressive sleep latency decrease of 34 minutes, or a ramelteon trial that finds a rather anaemic 9 minutes. The only fair way to compare ramelteon and melatonin is to run a head-to-head trial.

The only such trial that has ever been performed was performed on monkeys, and its results were contradicted by other monkey experiments. Also, it was run by the company that sells Ramelteon.

I think we may have enough evidence to conclude that Ramelteon is at least as effective as melatonin. There may even be some very tenuous evidence to suggest it is slightly more effective. But let me tell you a story.

One of my patients ran into the Ambien Walrus the other day and so, make a long story short, she needed a new sleeping pill. She was on a lot of drugs at the time and not all that healthy, and every drug I could think of, the pharmacist had some good reason why that would be a terrible idea in her case. Finally in desperation I remembered Ramelteon, which is safe as houses. Unfortunately Ramelteon is kind of new, and the pharmacy didn't have it.

"Okay," I said. "Why don't we just give her some melatonin? Some studies in monkeys suggest it *might* be slightly inferior to Ramelteon, but it's sure better than nothing."

Let's see if you are cynical enough to predict what happened next.

That's right. The hospital pharmacy, which carries thousands of drugs including bizarre experimental concoctions and super-expensive recombinant monstrosities, *didn't have melatonin*.

So do you want to know what the plan was, that the pharmacist and I came up with to treat my patient? I would take my lunch break, drive home, go into the cabinet in my bathroom, take the bottle of melatonin I had there, and bring it to the 500-something bed, multi-billion dollar hospital I work at.

This is why the story of Ramelteon scares me so much – not because it's a bad drug, because it isn't. But because one of the most basic and useful human hormones got completely excluded from medicine just because it didn't have a drug company to push it. And the only way it managed to worm its way back in was to have a pharmaceutial company spend a decade and several hundred million dollars to tweak its

chemical structure very slightly, patent it, and market it as a hot new drug at a 2000% markup.

I'm not knocking the pharmaceutical companies – they didn't do a think to suppress melatonin. All they did was notice that doctors were too dumb to use melatonin on their own and figure out a way around that problem.

And this is not an isolated incident. For example, on the rare occasions psychiatrists remember that folic acid exists at all they prescribe Deplin ($100/month, prescription only) instead of the *chemically identical* l-methylfolate ($5/month, over the counter).

While we're on the subject of melatonin, here are some Fun Melatonin Facts you may not have known (courtesy of *Melatonin and Melatonergic Drugs as Therapeutic Agents: Ramelteon and Agomelatine, the Two Most Promising Melatonin Receptor Agonists*):

— Melatonin's sleep promoting effects might be related to its ability to decrease core body temperature, which seems tantalizingly related to the finding that cooling caps are highly effective against insomnia.

— Smith-Magenis Syndrome is a rare genetic condition among whose effects are disruptions in the melatonin system. People with this syndrome wake at night and sleep during the day, meaning we can add this to porphyria, anemia, and rabies on the List Of Diseases That People With More Desire To Explain Away Ancient Folktales Than Sense Use As A Factual Basis For Vampirism.

— Many people use melatonin at night to try to hack their own circadian rhythms, but this is only mildly effective because they still have their own endogenous melatonin doing their own thing. The nuclear version of this strategy is to use

melatonin at night to increase melatonin levels and beta-blockers in the morning to decrease melatonin levels; the combination can give you almost complete control over your own circadian rhythm.

— Melatonin seems to play a role in fat metabolism and has been found to decrease weight gain associated with overfeeding in rats.

— Agomelatine is a melatonergic antidepressant that has been found to be approximately as effective as SSRIs with fewer side effects which is available in Europe. However, attempts to sell it in the USA were cut short when it failed to clearly differentiate from placebo in clinical trials (see: “found to be approximately as effective as SSRIs”)

— Melatonin appears to slow the growth of tumors, and a possible role as an adjuvant to classical chemotherapy drugs in cancer treatment is just one of the exciting areas of melatonin biology doctors are completely failing to explore.

In Defense of Psych Treatment for Attempted Suicide

A lot of the comments in my recent post on the implicit association test asked for a defense of why society should be hospitalizing suicidal people in the first place. If people have, after much thought, decided they prefer death to life, isn't that their right?

I am *extraordinarily* sympathetic to this position, which has been most eloquently defended by Sister Y of The View From Hell. Sister Y lists many harmful effects of suicide prohibition and many reasons why rational people might want to end their lives. She suggests a policy of legalizing fatal doses of barbituates for people who want them, allowing people tired of existence to leave the world without grisly suicide attempts that might leave them permanently injured or cause collateral damage to bystanders. I can't find her opinion on whether these should be provided on demand or whether you should have to undergo a psychiatric assessment first.

If she in fact believes the latter, then I think that position is defensible, and for professional reasons I won't publicly say anything further than that. But this post is to explain why it should require one *hell* of a psychiatric assessment and why the overwhelming majority of real-world suicide attempters would and should fail such an assessment.

Again, my point of disagreement is not on the ethics involved of letting some hypothetical perfect philosopher commit suicide – nor even on the fact that perhaps some cases genuinely are these perfect philosophers including Sister Y herself. I am trying to emphasize the practical point that in the real world, attempted suicides are rarely perfect philosophers

and almost always people who have made sudden, impulsive, and very bad decisions.

The greatest burden of suicide is of course to the friends and family of the person involved. But you don't have to be a Randian to think it's morally abominable to require someone in pain to continue living solely to please other people, so this post will focus solely on the welfare of the person involved.

What Does Youth Suicide Tell Us About Adult Suicide?

Start with the clearest case. About 4% of teenagers attempt suicide at some point (there are some much higher values from the CDC, which is usually pretty trustworthy, but some of the comments point out reasons why their estimates here are pretty hard to believe.)

I've gotten to observe some teenagers admitted to hospitals for attempted suicide. Some have incipient mental disorders that no one has noticed or considered treating. Some have unbearable home lives. Others have the standard litany of teenage problems – broke up with their boyfriend/girlfriend, bullied by the popular kids at school, got into a fight with their parents. Many have a combination of all three.

Some were the classic "cries for help" that were never meant to actually end in death, but others were entirely serious. A tragic few *intended* to take enough of an overdose to make Mom scared but not enough to actually kill them, but muddled their pharmacology in the most permanent possible way.

And I think most people agree that teenage suicide is terrible and requires treatment. Heck, most people won't even let teenagers make the decisions of whether or not to purchase alcohol, let along the decision to end their own lives. But I think aside from the inherent tragedy of teenage suicide it illuminates something about adult suicide as well.

These people have only the tiniest glimmer of knowledge about the likely happiness of their future lives. Most of their problems – the bullying by popular kids, the failed first relationship, even the awful families – are eminently wait-out-able. "It gets better" is not just for gay people (who, by the way, have a suicide rate up to 15x that of the straight population).

And yet teens attempt suicide at staggering rates.

There are certain depths of despair dark enough that the knowledge that the despair is completely temporary cannot penetrate them. It is this state, defined by the clouding of rationality by suffering, that I think most teenage suicides occur in.

And it would be very strange if this suddenly changed as soon as the victim hit eighteen.

Connection Between Suicides And Mental Disorder

It is generally reported that about 90% of suicides have some mental disorder. No, this isn't an artifact of psychiatrists assuming anyone who commits suicide *must* have a mental disorder – various half-decent methodologies have all converged around the same number, including a multitude of controlled studies (where psychiatrists evaluate a subject's mental status based on notes before knowing whether the person committed suicide) and prospective studies (where people only count as mentally disordered if they were diagnosed before the suicide occurred).

Sister Y has tried to poke holes in these statistics. First, she noted that the controlled studies showed 37% psych diseases even in the control population. But this number is probably correct – NIMH estimates that about 26% of people have mental disorders in a given year, and no doubt that number is

significantly higher among people who make good controls (ie are matched on demographic factors) for suicides. Second, she pointed out that the number included what she considered relatively "minor" disorders like alcohol dependence.

So first of all, alcohol dependence probably septuples your chance of committing suicide and something like 25% of suicides include alcohol. So I don't think it's unfair to include that in the list of how suicide is influenced by mental disorder.

But second of all, let me give totally anecdotal and probably unrepresentative examples of some other ways mental disorder can affect suicide.

As stereotypical as it sounds, the voices in people's heads do tell them to kill themselves a lot. Voices in people's heads are *huge jerks* and occasionally people will do what they say just to make them shut up. The tragedy here is that antipsychotic drugs are pretty good at dealing with this if people can just get access to them. Among schizophrenia patients (the group most commonly identified with these sorts of symptoms), almost half attempt suicide and 10% complete it. Since schizophrenics make up 1% of the general population, that's a non-negligible fraction of total suicides.

You know what's an even less fun form of psychosis? Psychotic depression. This is where people get so depressed they start hallucinating about how horrible they are. I will never forget the patient who stopped eating because she believed her digestive system was rotting away and infested with maggots. And a lot of the time these people's self-hatred reaches completely bizarre proportions in which they will confess to causing the Holocaust or the 9-11 attacks just because *it seems like the sort of thing someone as horrible as them might do*. If you believe you caused the Holocaust, this

seems like a pretty good reason to kill yourself in the name of justice, and sadly this is what many of these people do. And again, this is tragic because psychiatry is actually not so bad at dealing with this kind of over-the-top depression (the rotting-intestines woman became much better after a short course of electroconvulsive therapy, but some people will get better just on medications).

Borderline Personality Disorder is another common cause of suicides. It intensifies emotions so that anyone so much as making a mildly critical remark makes you think everyone will hate you forever and you deserve to die. And then six hours later someone smiles at you and you feel like the world is perfect and beautiful. But if you commit suicide at one of the low points, then that's *it*. And Borderline Personality Disorder, again, is *sorta* amenable to therapy, and even without therapy half the time it just *goes away* after a few years to a decade.

Alcohol and drug abuse is another big one. Some of it is that abusers have worse lives – poor health, financial issues, more likely to have trouble at work. But a big part of it is just *lowered inhibition*. If a sober person is walking on a bridge after some life crisis, they might have fleeting thoughts of jumping but suppress them after thinking of the future. If a drunk person is walking on a bridge after some life crisis, the frontal lobes that would normally suppress those urges are partly out of commission.

And then there's depression. I'm trying not to make a big deal about it because everyone associates suicide and depression when in fact the correlation is no higher than many other mental illnesses (although the greater number of depressed people does make absolute numbers higher). I guess all I'll say here beyond what everyone already knows is that Major Depressive Disorder (classic depression) is an intermittent

disease. The average depressive episode lasts less than six months, and the average person with MDD has only four depressive episodes in their lifetime (these numbers are even better if you're on medication, which many depressed people fail to be). There's a thing called dysthymia, which is like having depression all the time, but it is thankfully less common and less severe and not where most suicides are coming from.

I am certain that six months *feels* like an eternity if you are depressed. And no doubt knowing that you're going to have to deal with the same thing a few more times in your life (ALTHOUGH SERIOUSLY, MEDICATION DOES HELP WITH THIS) must also be, well, depressing. But the average depressive suicide is not a Perfect Philosopher who has calculated, while healthy, that the possibility of another six month depressive episode is too much to bear.

The average depressive suicide is someone in the middle of one of their episodes who, like the teenagers above, is in the place so dark that they've forgotten the existence of hope. They're somewhere so dark that "this will probably go away in a couple of months" has no meaning. Somewhere so dark that one of the main side effects of *effective* antidepressant drugs is suicide, because a few weeks after starting the patient finally has enough energy to go kill themselves, but doesn't consider waiting a month or so for the drug to take full effect.

I want to end this section with a study – small, but encouraging – that cognitive-behavioral therapy (aka That One Type Of Psychotherapy That Sometimes Works) reduces suicide 50% in at-risk populations. Think about that. What percent of suicides do you think haven't had cognitive-behavioral therapy? 80%? 90%? Whatever that percent is, half

of them *would have been fine* if they had just had access to a good psychologist.

Empirically, Suicides Regret It

People who commit suicide can't change their minds. But attempted suicides can and do, and we can analyze these changes both in their actions and in their words.

In terms of revealed preferences, most people who are prevented from completing their suicide do not go on to kill themselves. Sister Y critiques a study saying only 4% later go on to kill themselves, and offers as counterpoint a study she prefers claiming 13% do (she finds a way to round up to 19%). I have also heard 10%, although I can't remember where. Do you know what the numbers 4%, 10%, 13%, and 19% all have in common? Yes. They are all significantly less than 50%.

It is somewhat harder to find good studies on what percent *attempt* suicide again. By eyeballing some other statistics and trying to fit them together, I believe it is greater than 25% but less than 50%. One textbook whose studies I have not been able to verify says that 30% of untreated and 15% of treated suicide attempters try again. 15% and 30% are also among the many numbers that are less than 50%.

And keep in mind what these data *don't* show. They don't show that the 25-50% who try again have lives so constantly miserable that they continue wanting to die. Remember that intermittent depression from before? Imagine a world in which depressive episodes last one day each, and people only have two of them in their lives. Other than those two days, they live happy lives and are grateful to be alive. Doesn't matter. This pattern would *still* be consistent with 25-50% of attempted suicides making repeated attempts, if that second day of depression was bad enough

(out of fairness I should mention this data *also* doesn't show that 50-75% of people get over their suicidality; it's consistent with them just being tired of suicide attempts not working and settling for continued existence. I guess what I'm saying is that the data don't prove very much)

So moving from boring data to the much-more-fun domain of anecdote, a surprising number of suicide attempters change their mind *during* the suicide attempt. One particularly famous case is that of Kevin Baldwin, who survived jumping off of America's favorite suicide spot. He says that while still in the air "I instantly realized that everything in my life that I'd thought was unfixable was totally fixable—except for having just jumped."

Most realizations are slightly less dramatic, but my work in a psych ER taught me that many 9-1-1 calls about suicides are from the victims themselves. I remember one patient, a typical case, who overdosed on pills. As she lay on the ground starting to feel sick, she thought about her problems a little more deeply, thought about how her family would feel, and decided she preferred to live. She called 9-1-1, they sent an ambulance over, and the hospital managed to keep her alive until the drugs passed out of her body. This is quite common. It also contradicts one of Sister Y's strongest arguments – that the reason many people avoid suicide is out of fear of making the attempt. A non-negligible number of people who have already made the attempt and just have to sit back and day find themselves changing their minds and actively working to save their own lives.

But most of the stories I can generate from my personal experience are nothing more dramatic. It's people who were found by their parents or partners or friends, dragged kicking and screaming to the hospital, treated for a couple of days, and

by Day 3 they're saying oh my god I made a horrible mistake I can't believe what almost happened.

And I know what the response will be – that of *course* they'd say that to their psychiatrists, they're trying to get judged Officially Sane so they can get discharged and maybe try again. I accept that as a possibility, but since this whole section is about totally useless anecdotal data, let me just say I don't feel like that was what was happening. I met people who were going out of their way to look for and thank their psychiatrist when he was busy in his office after the discharge papers had already been signed and they were on their way out. One time I met a patient at the bus stop a few days after she had been discharged, and she asked me to thank my boss and the rest of the team for what must have been the umpteenth time.

Finally, I have some personal friends who have attempted suicide. In every case I am *incredibly* glad they remain alive, and more importantly, usually they are as well. And I know there's social pressure here – that psychiatrists aren't the only ones you have a vested interest in appearing cheerful to – but some are very close to me indeed and I do not believe they would lie about something this important.

Psychiatric Care Probably Helps

One of the most common objections to sending people who attempt suicide to psychiatric hospitals is that it is a terrible punishment, that we are essentially locking up and drugging and torturing people whose lives are already apparently pretty bad.

But mental hospitals for people who attempt suicide (actually almost always just the psychiatry floor of a regular hospital) are *not like that one in One Flew Over The Cuckoo's Nest*. I can't repeat that enough. I know that as a psychiatrist-in-

training I have no credibility on this issue, so take it from a former psychiatric patient. Your problems are much more likely to be along the lines of a terrible selection of books in the ward library than torture by sadistic nurses (I do not deny the latter occurs, just as some schools have torture by sadistic teachers, but it is extremely rare and nowhere near for example what goes on in nursing homes).

According to the CDC, the average length of stay in a mental health ward is one week (this brief by an organization I've never heard of says 8 days). That includes catatonic people and people who have long animated conversations with the Devil, so the average suicidal person isn't going to be the one bringing up that average.

In practice I have a pretty good guess for the *exact* length of stay the average suicidal person without associated mental disorders will experience, and that is 72 hours. That's the maximum amount of time a hospital can legally commit someone against their will. After that they have to get a court order allowing them to hold the patient longer, and this requires swearing that the patient is mentally incompetent to make their own decisions, and most doctors will not do this without reason.

But if you don't trust doctors' benevolence, at least trust their self-interest: it takes a lot of paperwork, it requires them to go all the way to a courthouse, and the hospital management is going to be breathing down their back the whole time about how they could *really* use an extra bed on Ward 4 and of course we would *never* pressure you to discharge any patients before they're better, but seriously, have a bed open on Ward 4 by tomorrow. Trust me, doctors are not plotting to keep people in the hospital longer than necessary. If you like conspiracy

theories, the opposite conspiracy is a much bigger cause for concern.

That's usually just enough time to evaluate the patient for mental disease, start them on some medication, and refer them to an outpatient psychologist and/or psychiatrist. One hospital I worked at kept (mostly willing) people in a little longer to see if the drugs actually took effect, but that was a luxury they could only afford because they were a rich academic institution.

But the thing is, *this really helps*. If 90% of people committing suicide have some associated mental disease, and mental diseases can dectuple your risk of committing suicide, then connecting these people – many of whom have never interacted with the mental health system before – with someone who can help them (or even with a Prozac prescription) can be a really, really big deal.

I mentioned before that one specific form of therapy can decrease future suicide rates 50%. That was in a study where *both* groups were getting the recommended psychiatric drugs. Another study I cited above said that "psychiatric treatment" (whatever that means; I bet it didn't include the CBT from the last study and so they're cumulative) can also cut future suicide rates in half. There are more specific studies on the anti-suicide effect of each individual drug – lithium is an example of a particularly good one.

(fun fact which there is a small chance I will devote my life to studying: even areas with slightly higher trace amounts of lithium in the *water supply* have lower suicide risk.)

And even if you're one of the depressingly high number of people who throw away their prescription and never show up to their psychiatrist, you know what? You've been stuck in a

big building with lots of people watching you for the three days or so immediately after whatever horrible event made you become suicidal in the first place. Drugs in your system? Now you're clean. Angry at a family member? Maybe you're less angry now. Upset over a breakup? Maybe you've had a chance to think about it a little more.

I am very reluctant to get into in what situations I believe suicide is acceptable. I am scared that one day my future employers will read this post. Or worse, a future patient will read it and start arguing "You said suicide was acceptable if A or B, so I did those things". So all I will say is that I wish Sister Y and those like her maximum utility however they define their utility function. But anyone considering suicide who has thought about it less than she has or lacks her philosophical acumen should consider getting professional help (or even non-professional help) or at least meditate long and hard on that cliche about "a permanent solution to a temporary problem".

EDIT: Since people are missing something I said like *a thousand times* in the post itself, I'll put it down here in bold. **I am not claiming that suicide is never rational and that all suicides are stupid and impulsive, or that no one can ever legitimately want to die. I am saying those people make up a very small portion of suicides, and that the typical case is people who do it impulsively or in a state where they lack full decision-making capacity. And that the psychiatric system can be of huge help to this latter group, and that helping the former group is a different question which I do not want to talk about publicly for professional reasons.**

Who By Very Slow Decay

[*Trigger warning: Death, pain, suffering, sadness*]

I.

Some people, having completed the traditional forms of empty speculation – "What do you want to be when you grow up?", "If you could bang any celebrity who would it be?" – turn to "What will you say as your last words?"

Sounds like a valid question. You can go out with a wisecrack, like Oscar Wilde ("Either this wallpaper goes or I do"). Or with piety and humility, like Jesus ("Into thy hands, o Father, I commend my spirit.") Or burning with defiance, like Karl Marx ("Last words are for fools who haven't said enough.")

Well, this is an atheist/skeptic blog, so let me do my job of puncturing all your pleasant dreams. You'll probably never become an astronaut. You're not going to bang Emma Watson. And your last words will probably be something like "mmmrrrgggg graaaaaaaaaaHAAACK!"

I guess I always pictured dying as – unless you got hit by a truck or something – a bittersweet and strangely beautiful process. You'd grow older and weaker and gradually get some disease and feel your time was upon you. You'd be in a nice big bed at home with all your friends and family gathered around. You'd gradually feel the darkness closing in. You'd tell them all how much you loved them, there would be tears, you would say something witty or pious or defiant, and then you would close your eyes and drift away into a dreamless sleep.

And I think this happens sometimes. For all I know, maybe it happens quite a lot. If it does, I never see these people. They

very wisely stay far away from hospitals and the medical system in general. I see the other kind of people.

If you are like the patients I see dying, then here is how you will go.

You will grow old. When you were young, you would go to institutions and gradually gather letters after your name: BA, MD, PhD. Now that you are old, you do the same thing, but they are different institutions and different letters. Your doctors will introduce you to their colleagues as "Mary Smith, COPD, PVD, ESRD, IDDM". With each set of letters comes another decrease in quality of life.

At first these sacrifices will be minor. The COPD means you have to breathe from an oxygen tank you carry around wherever you go. The PVD will prevent you from walking more than a few feet at a time. The ESRD will require three hours dialysis in a hospital or outpatient dialysis center three times a week. The IDDM will require insulin shots after every meal. Not fun, but hardly inconsistent with a life worth living.

Eventually these will add up beyond your ability to manage them on your own, and you will be sent off to a nursing home. This will seem like a reasonable enough idea, and sometimes it goes well. Other times it gives you freedom to develop a completely new set of morbidities totally unconstrained by what a person in any other situation could possibly be expected to survive.

You will become bedridden, unable to walk or even to turn yourself over. You will become completely dependent on nurse assistants to intermittently shift your position to avoid pressure ulcers. When they inevitably slip up, your skin develops huge incurable sores that can sometimes erode all the way to the bone, and which are perpetually infected with foul-

smelling bacteria. Your limbs will become practically vestigial organs, like the appendix, and when your vascular disease gets too bad, one or more will be amputated, sacrifices to save the host. Urinary and fecal continence disappear somewhere in the process, so you're either connected to catheters or else spend a while every day lying in a puddle of your own wastes until the nurses can help you out. The digestive system isn't too happy either by this point, so you can either have a tube plugged directly into your stomach or just skip the middleman and have an IV line feeding nutrients into your bloodstream.

Somewhere in the process your mind very quietly and without fanfare gives up the ghost. It starts with forgetting a couple of little things, and progresses until you have no idea what's going on ever. In medical jargon, healthy people are "alert and oriented x 3", which means oriented to person (you know your name), oriented to time (you know what day/month/year it is), and oriented to place (you know you're in a hospital). My patients who have the sorts of issues I mentioned in the last paragraph are generally alert and oriented x0. They don't remember their own names, they don't know where they are or what they're doing there, and they think it's the 1930s or the 1950s or don't even have a concept of years at all. When you're alert and oriented x0, the world becomes this terrifying place where you are stuck in some kind of bed and can't move and people are sticking you with very large needles and forcing tubes down your throat and you have no idea why or what's going on.

So of course you start screaming and trying to attack people and trying to pull the tubes and IV lines out. Every morning when I come in to work I have to check the nurses' notes for what happened the previous night, and every morning a couple of my patients have tried to pull all of their tubes and lines out.

If it's especially bad they try to attack the staff, and although the extremely elderly are *really* bad at attacking people this is nevertheless Unacceptable Behavior and they have to be restrained ie tied down to the bed. A presumably more humane alternative sometimes used instead or in addition is to just drug you up on all of those old-timey psychiatric medications that actual psychiatrists don't use anymore because of their bad reputation.

After a while of this, your doctors will call a meeting with your family and very gingerly raise the possibility of going to "comfort care only", which means they disconnect the machines and stop the treatments and put you on painkillers so that you die peacefully. Your family will start yelling at the doctors, asking how the hell these quacks were ever allowed to practice when for God's sake they're trying to kill off Grandma just so they can avoid doing a tiny bit of work. They will demand the doctors find some kind of complicated surgery that will fix all your problems, add on new pills to the thirteen you're already being force-fed every day, call in the most expensive consultants from Europe, figure out some extraordinary effort that can keep you living another few days.

(then these people will go home and log onto the Internet and yell at cryonics advocates for being selfish for wanting to live longer. Don't those stupid cryonicists realize all that money could be spent on charity, instead of chasing after fantastically unlikely chances?)

Robin Hanson sometimes writes about how health care is a form of signaling, trying to spend money to show you care about someone else. I think he's wrong in the general case – most people pay their own health insurance – but I think he's spot on in the case of families caring for their elderly relatives. The hospital lawyer mentioned during orientation that it never

fails that the family members who live in the area and have spent lots of time with their mother/father/grandparent over the past few years are willing to let them go, but someone from 2000 miles away flies in at the last second and makes ostentatious demands that EVERYTHING POSSIBLE must be done for the patient.

Your doctors will nod their heads and tell your family they respect their wishes. It will be a lie. Oh, sure, they will *carry out* the family's wishes, in terms of continuing to provide the care. But *respect*? In the cafeteria at lunch, they will – despite medical confidentiality laws that totally prohibit this – compare stories of the most ridiculous families. "I have a blind 90 year old patient with stage 4 lung cancer with brain mets and no kidney function, and the family is *demanding* I enroll her in a clinical trial from Sri Lanka." "Oh, that's nothing. *I* have a patient who can't walk or speak who's breathing from a ventilator and has anoxic brain injury, and the family is insisting I try to get him a liver transplant."

Every day, your doctors will meet with your family another time, and eventually, as your condition worsens and your family has more time to be hit on the head with a big club marked 'REALITY', they will start to relent. Finally, they will allow your doctors to take you off of the machines, and you will be transferred to Palliative Care, whose job I do not envy even though *every single palliative care doctor I have ever met is relentlessly cheerful and upbeat and this is a total mystery to me*.

And you will die, but not quickly. It takes time for the heart to give up, for the lungs to fill with water and stop breathing, for the toxic wastes to build up. It is generally considered wise for the patient to be on epic doses of morphine throughout the process, both to spare them the inevitable pain as their disease

takes their course and to spare their family from having to watch them.

…not that they always do. It can take anywhere from a day to several weeks for someone to die. Sometimes your family wants to wait at the bedside for a week. But a lot of the time they have work and things to do. Maybe they live thousands of miles away. You haven't recognized them in years, you haven't spoken a coherent word in months, and even if for some reason your brain chose this moment to recover lucidity you're on enough morphine to be *well* inside the borders of la-la-land. A lot of families, faced with the prospect of missing work and school to sit by what's basically a living corpse day in and day out for weeks just to watch it turn into a non-living corpse, politely decline. I absolutely 100% cannot blame them.

There is a national volunteer program called No One Dies Alone. Nice people from the community go into hospitals to spend time with dying people who don't have anyone else there for them. It makes me happy that this program exists.

Nevertheless, this is the way many of my patients die. Old, limbless, bedridden, ulcerated, in a puddle of waste, gasping for breath, loopy on morphine, hopelessly demented, in a sterile hospital room with someone from a volunteer program who just met them sitting by their bed.

And let me just emphasize again, not everyone dies this way. I am hugely selection biased by my position in a hospital. But enough people die this way. I'm in a small community. There can't be too many deaths here. Of the ones there are, I see a lot of them. And they're not pretty.

*[**EDIT:** Just looked up statistics. Only about a quarter of old people die at home. The rest are split between hospitals (disproportionately ICUs), nursing homes, and hospices.]*

II.

Hospital poetry is notoriously bad.

I mean, practically all modern poetry is bad. Modern poetry by complete amateurs could be expected to be even worse. But hospital poetry is in a league all of its own as far as badness goes.

When I search "hospital poetry", Google brings up examples like the following:

> Pain… searing
> Belly… throbbing
> There is no baby.
> There will be no baby.
> Endometriosis.

I feel bad making fun of it, because it is clearly heartfelt. This is part of the problem with hospital poetry. It is very heartfelt, whereas I think most popular poetry comes from people who have strong emotions but also some distance from them and a little bit of post-processing. And unfortunately doctors, who are on this decades-long quest to prove they are actual people with real feelings and not just arrogant robot-like people in white coats who know a very large number of facts about thyroiditis, just eat this sort of thing up.

But I'm not really complaining about those sorts of endometriosis poems. The ones I'm really complaining about are worse. The epitome of the genre I can't find on Google, because it was presented as some kind of event at the hospital where I trained in Ireland. I don't remember it, but let me just make up some doggerel approximately faithful to the spirit of the original:

When my doctor told me that I had cancer
I knew that despair was not the answer
It felt like the darkness was closing in
But to give up would have been a sin
Everyone here helped me so much
And nothing is like a helping hand's touch
Thanks, Dr. Connell, and everyone in Cork
I really appreciate all your hard work

Doctors and nurses eat this kind of thing up and put it on shiny plaques that go on the walls of the hospital. (I suggest a wall near the gastroenterology unit, to expedite care for people who start vomiting.)

Wittgenstein said that "if anyone ever wrote a book of ethics, that really was a book of ethics, it would destroy all the other books in the world with a bang." I'm not really sure what he meant. But if anyone ever wrote a book of hospital poetry, that really was a book of hospital poetry…well, I don't know what would happen, but I bet it would be loud and angry, and that it wouldn't be put on shiny plaques on anybody's walls, except maybe the same people who hang Hieronymous Bosch paintings on their walls.

Wait, am I calling hospitals hellish? Sure am. It has nothing to do with the decor, which has actually gotten much nicer in your newer hospitals until it's hard to tell them apart from a stylish office building. It's nothing to do with the staff, either – most doctors and some nurses seem pretty happy and trade banter around the water coolers like everyone else. It's mostly the screams.

The screams are coming about 33% from the confused demented old people I mentioned, 33% from people having minor procedures performed without anaesthetics for one or

another good reason, and 33% from people who just have very painful diseases (plus 1% from me sitting in the break room looking up examples of hospital poetry for this post). They run the gamut of human screams. There are wordless shrieks. There are some angry screams, like "$#%! YOU GET ME OUT OF HERE!". There are a lot of people screaming "SOMEBODY HELP ME!" And there are some religious screams, like "OH GOD!" or "JESUS HELP ME!" or "CHRIST NO!".

When I first started working in hospitals, I would not only inevitably run over to these screams, but I would feel contempt and anger at the rest of the hospital staff who would just continue their daily routine. I soon learned better. Not only would I be unable to do anything – I can't single-handedly cure their painful illness, or make their procedure go any faster, or explain to them that the year is 2013 and they're no longer on their childhood farm in Oklahoma – but as soon as they saw me I would be the one they started screaming at and expecting to save them. The bystander effect, my last defense, disappeared. Sometimes I would make a stand by asking the nurse to increase their pain medication or something, and be politely told all the reasons why that was a bad idea from a medical perspective (pain medication has lots of side effects which doctors monitor carefully). In the end I would just slink out of the room, wishing I had never come in.

So the constant screams being completely ignored by a bunch of happy people going through their day is pretty hellish. But there's also the bodies. Usually we are able to avoid thinking about people as bodies except to briefly note that certain people like Emma Watson are really hot. In a hospital, this filter disappears. Some people have gigantic swollen legs the size of your waist. Others have huge ulcerated sores all over.

Still others have skin covered with the sorts of bacterial colonies you usually only see on a petri dish. And body sizes range from so thin that you can see their organs bulging out of their skin and use them as a grisly impromptu anatomy lesson, to so morbidly obese that you have to search through the fat folds to find body part you're looking for.

The senses are under constant assault. Smell is the worst. There are some people who can identify different infections by smell. *Pseudomonas aeruginosa* is supposed to smell fruity. *Gardnerella* is supposed to smell fishy. *Clostridium* is supposed to smell like the worst thing you can possibly imagine, if it were then covered in feces and left to rot on a warm summer day.

But the other senses get their time too. The sight is vexed by flashing call lights. And the hearing is battered with incessant beeping from IV lines which have hard-coded alarms to alert doctors of critically important events such as "Look at me! I am an IV line!" The end result is something it would take a first-rate poet to describe. I'm tempted to nominate Oscar Wilde. He did a good job on prisons in *Ballad of Reading Gaol*, and I feel like the skill would transfer:

> He does not rise in piteous haste
> To put on convict-clothes,
> While some coarse-mouthed doctor gloats,
> and notes each new and nerve-twitched pose,
> Fingering a watch whose little ticks
> Are like horrible hammer-blows […]
>
> He does not stare upon the air
> Through a little roof of glass;
> He does not pray with lips of clay
> For his agony to pass;

Nor feel upon his shuddering cheek
The kiss of Caiaphas.

But after some more thought, I think I'm going to go with Wilfred Owen:

If in some smothering dreams you too could pace
Behind the wagon that we flung him in,
And watch the white eyes writhing in his face,
His hanging face, like a devil's sack of sin;
If you could hear, at every jolt, the blood
Come gargling from the froth-corrupted lungs,
Obscene as cancer, bitter as the cud
Of vile, incurable sores on innocent tongues […]

Or better yet, if Oscar Wilde's muse when he was writing *Reading Gaol* were to bear Wilfred Owen's children, then those kids would be competent to write hospital poetry that was actually hospital poetry.

Dante would also be an acceptable choice.

III.

You may have read the excellent article How Doctors Die. If you haven't, do it now. It says that most doctors, knowing everything I've just mentioned above, choose to die quickly and with very limited engagement with the health system.

I (and the doctors in my family whom I've asked) am pretty much like the doctors in the article. If I get a terminal disease, I want to wring what I can out of the few months of life I have left and totally avoid any surgery, chemotherapy, amputations, ventilators, and the like. It would be a clean death. It would be okay.

My big fear, though, is that I *won't* get a terminal disease.

If I just start accumulating damage, growing more and more bedridden and demented and pain-riddling until I want out – well, there won't *be* a way out. If there's not some very specific life-saving treatment that can be withdrawn, I'm stuck above ground, not just in the "unless I want to risk the danger and shame of suicide" way I am now, but – if I'm too debilitated to access means of suicide on my own – in an absolute way.

Even if my doctors and nurses and caretakers are sympathetic, my only legal option, without exposing *them* to jail time, is to starve myself to death – something both painful and difficult, and itself not really the way I want to go.

I was sitting in an ICU room yesterday where a patient's body had just been brought out after their death. My attending was taking care of the paperwork in the other room, and I was sitting there reflecting, and I started thinking about what it would be like to die in that room. There was a big window, and it was a sunny day, and although I mostly had a spectacular view of the hospital parking lot, a bit further in the distance I could see a park full of really big trees. And I knew that if I were dying in that room my last thought would be that I wanted to be outside.

I think if I were very debilitated and knew I would die soon, I would want to go to that park or one like it on a very sunny day, surround myself with my friends and family, say some last words, and give myself an injection of potassium chloride.

(this originally read "morphine", but just today the palliative care doctor at my hospital gave an impassioned lecture about how people need to stop auto-associating morphine with euthanasia, because it makes it really hard for him to offer

morphine painkillers to patients who need them without them freaking out. So potassium chloride it is.)

This will never happen. Or if it did, it would be some kind of huge scandal, and whoever gave me the potassium chloride would be fired or something. But the people dying demented and hopeless connected to half a dozen tubes in ICU rooms aren't considered scandals by anybody. That's just "the natural way of things".

I work in a Catholic hospital. People here say the phrase "culture of life" a lot, as in "we need to cultivate a culture of life." They say it almost as often as they say "patient-centered". At my hospital orientation, a whole bunch of nuns and executives and people like that got up and told us how we had to do our part to "cultivate a culture of life."

And now every time I hear that phrase I want to scream. 21st century American hospitals do not need to "cultivate a culture of life". We have enough life. We have life up the wazoo. We have more life than we know what to do with. We have life far beyond the point where it becomes a sick caricature of itself. We prolong life until it becomes a sickness, an abomination, a miserable and pathetic flight from death that saps out and mocks everything that made life desirable in the first place. 21st century American hospitals need to cultivate a culture of life the same way that Newcastle needs to cultivate a culture of coal, the same way a man who is burning to death needs to cultivate a culture of fire.

And so every time I hear that phrase I want to scream, or if I cannot scream, to find some book of hospital poetry that really is a book of hospital poetry and shove it at them, make them read it until they understand.

There is no such book, so I hope it will be acceptable if I just rip off of Wilfred Owen directly:

> If in some smothering dreams you too could pace
> Behind the gurney that we flung him in,
> And watch the white eyes writhing in his face,
> His hanging face, like a devil's sack of sin;
> If you could hear, at every jolt, the blood
> Come gargling from the froth-corrupted lungs,
> Obscene with cancer, bitter with the cud
> Of vile, incurable sores on innocent tongues
> My friend, you would not so pontificate
> To reasoners beset by moral strife
> The old lie: we must try to cultivate
> A culture of life.

Medicine, As Not Seen on TV

Since I was twelve years old, my life has taken place in a series of Four Year Intervals.

Four years of high school. Four years of college. Four years of medical school. Four years of residency. Four times four, nice and symbolic.

This comes to mind now because I finished my first year of residency today.

I went into it raised on a steady diet of medical TV dramas like *Scrubs* and *House*, the legends passed down by other doctors in my family, and the ideas inculcated into me in medical school. It turned out to be nothing like any of those.

I've written a few posts about my experiences at work: The Hospital Orientation, I Aten't Dead, Who By Very Slow Decay, and Evening Doc. I've tried to avoid writing anything more specific in order to protect patient confidentiality and *my* confidentiality.

But I thought this would be a good time to record – for my future self as much as for anyone else – what surprised me in my first year of medical practice.

To start with, forget about diagnostic mysteries. If you've ever seen *House* or anything else remotely like it, you imagine doctors as constantly presented with weird and wonderful symptoms, then racing against the clock to figure out what rare and deadly disease it is.

In real life, patients are more like the elderly lady I got last month. She had three hospital admissions for urinary tract infections in the past two years. Now she comes in with

urinary symptoms. Before I even know the patient exists, the emergency room doctor has run a urine test which reveals that it's a urinary tract infection. He has helpfully started her on the correct antibiotic for urinary tract infections. WHAT COULD THIS DIAGNOSTIC MYSTERY POSSIBLY BE?

Yeah, it was a urinary tract infection.

Or the guy who comes in shaking and sweating. I ask him what happened. He said he has been drinking alcohol for thirty years, and two days ago he tried to stop cold turkey. Have you ever had these sorts of symptoms before? Yes, every time I go off alcohol I get them. Does anything relieve the symptoms? Yes, drinking more alcohol. SOMEBODY PAGE DOCTOR HOUSE TO FIGURE OUT WHAT'S GOING ON?

Yeah, it was alcohol withdrawal.

Not all the patients I got were like this. But probably ninety-five percent of them were. Most people come into hospital for flare-ups of chronic problems they have had for, at minimum, ten years. Most of the time they have been to their primary care doctor first, who has made the diagnosis and sent the patient to the hospital for treatment. Or if not, they go to the emergency room, where the emergency room doctors do the same standard blood test they do on everybody and which usually gives you a really good idea what's up. Oh, you're feeling sick and tired and thirsty and nauseous? Hmm, your blood glucose is five hundred. Are you a diabetic? Did you take your insulin? Why didn't you take your insulin? "Being on vacation" is not a good reason to stop taking your insulin! Do you promise to take your insulin in the future? Okay, well let's admit you to the hospital and send you to Dr. Alexander so he can clear up this massive medical mystery we have on our hands.

But okay, five percent of cases we're not entirely sure what's going on. *Now* we can page Dr. House, right?

Wellll, in reality we "stabilize" them. A lot of the time "stabilize" means "put them in a bed and give them IV fluids and they get better on their own". Sometimes the problem looks vaguely infectious and so we give empiric antibiotics, where empiric means "let's give them an antibiotic that works for lots of stuff, and maybe it'll work for this". Sometimes the problem looks vaguely autoimmune and we give them steroids.

It's pretty funny, because in medical school you spend a *lot* of time learning about maybe two dozen very rare autoimmune diseases, and how to differentiate Wegner's granulomatosis from Takayasu arteritis, and the very subtle differences in the aetiology of each. And in real life, my attending says "Huh, this looks vaguely autoimmune, let's throw steroids at it." And it always works.

Now I understand that when the patient leaves hospital, they go to a rheumatologist or other specialist, and the specialist probably does lots of complicated tests and then comes up with a treatment regimen perfectly suited to that patient. But at the level I'm working at, it's more "Hey, it responded to steroids! I guess it really *was* autoimmune! Or maybe the patient just got better on her own. Or something. Anyway, who cares, patient's better, let's discharge before something goes wrong."

Because something else always goes wrong. You may be wondering: if doctors don't spend their time solving diagnostic mysteries, what *do* they do in all those long hours they work? The answer is: deal with the avalanche of disasters that

inevitably begin the second a patient walks through the door into a hospital.

I want to make it very clear I'm not criticizing my own hospital here. They make an *amazing* effort to do everything possible to avoid dangerous complications. All the hospitals I've worked at do. And all of them are death-traps. God just has a particular hatred for hospital patients, which He expresses by inflicting random diseases upon them for so long as they make the mistake of staying within the four walls and ceiling of a hospital building.

Like, you can be a perfectly healthy person, who lives forty years without anything worse than a sniffle. And then one day you're playing sports, and you break your leg and you think "What's the worst that can happen, I'll spend a day or two in the hospital?" and by the time you come out you've got two artificial legs and a transplanted kidney and a rare bunyavirus from the African tropics and you have to inject yourself with insulin every three hours or else you die.

There are some good reasons for this. Obviously hospitals are full of sick people which means the potential for contagious infectious is high. People in hospitals are always getting lines stuck into them and surgeries performed and otherwise having foreign objects stuck in the body, and of course that's a risk factor for all kinds of stuff. People in hospitals are often taking medications, which often have side effects. People in hospitals are often having tests, which sometimes involve injecting large amounts of radioactive material into the body and hoping it doesn't fry anything important.

Then there are reasons you never expect until someone teaches you about them. If you don't move your legs enough – maybe because you're lying in a hospital bed all day – the blood in

your legs settles and clots, and then the blood clots travel to your lungs, and then you can't get any oxygen and potentially die. If you don't fidget enough – maybe because you're lying in a hospital bed unconscious – the constant pressure on a single patch of skin produces an ulcer, which gets infected and you potentially die. If you take five different recreational drugs every day, and your dealer doesn't visit you in the hospital, then you go into withdrawal, and if you don't want to admit what's going on to your doctor maybe they miss it and – yeah, you potentially die.

But probably the biggest reason – and one you never think of – is that the hospital is where they're finally doing tests on you, which means all those diseases that were lying dormant before and which you put down to normal old age finally get detected. You come in for a kidney stone, but your doctor does a blood test and finds you have diabetes. Also your calcium is a little off, we're going to need to give you calcium pills and set up an appointment to get your parathyroid checked. And also when they did the CT of the kidneys they found a suspicious-looking mass in the colon, so you're going to have to get that checked out. Uh, the gastroenterologist pulled the joystick controlling the colonoscope a little too hard and now you have a perforated colon, you need surgery. Uh, the surgeon put on her gloves the wrong way, now the surgical site is infected, guess you need antibiotics. Uh, guess you're allergic to that antibiotic, let's use a different one. Wow, allergic to four antibiotics in a row, guess this isn't your day!

While Dr. House is diagnosing Chikungunya fever, the rest of us are treating the person who came in with a nosebleed (final diagnosis: blew nose too hard) but now has a DVT, hyperkalaemia, Sundowner's syndrome, and a line infection.

Well, sort of treating.

John Searle came up with this really interesting philosophy-of-consciousness thought experiment. Suppose that a man were put in a room with a bunch of books, each of which contained a set of rules about Chinese characters. Sometimes, a paper with Chinese characters would come in through a slot in the door. The man would apply the rules in his book, which told him to write certain Chinese characters if certain conditions about the characters on the paper held true, and slip the output back through the slot in the door. The man does this faithfully, although he doesn't know any Chinese and has no idea what any of it is saying.

On the other side of the door is a Chinese person. In her mind, she's writing questions to the man, and he is responding back in fluent Chinese. She thinks they're having a very productive conversation, and is starting to get a crush on him.

And the question is, in what sense can the man in the room be said to "understand" Chinese? If the answer is "not at all", then in what sense can the brain – which presumably takes inputs from the environment, applies certain algorithms to them, and then sends forth appropriate outputs – be said to understand anything?

Daniel Dennett and various other materialist philosophers have a response to this challenge, which is that the man does not understand Chinese, but the man, his books, and the room can be conceptualized as an emergent system that does possess the property of Chinese-understanding and which may or may not be conscious.

I bring this up, because I understand what's going on with patient care about as well as the man understands Chinese. I feel like maybe the hospital is an emergent system that has the

property of patient-healing, but I'd be surprised if any one part of it does.

Suppose I see an unusual result on my patient. I don't know what it means, so I mention it to a specialist. The specialist, who doesn't know anything about the patient beyond what I've told him, says to order a technetium scan. He has no idea what a technetium scan is or how it is performed, except that it's the proper thing to do in this situation. A nurse is called to bring the patient to the scanner, but has no idea why. The scanning technician, who has only a vague idea why the scan is being done, does the scan and spits out a number, which ends up with me. I bring it to the specialist, who gives me a diagnosis and tells me to ask another specialist what the right medicine for that is. I ask the other specialist – who has only the sketchiest idea of the events leading up to the diagnosis – about the correct medicine, and she gives me a name and tells me to ask the pharmacist how to dose it. The pharmacist – who has only the vague outline of an idea who the patient is, what test he got, or what the diagnosis is – doses the medication. Then a nurse, who has no idea about any of this, gives the medication to the patient. Somehow, the system works and the patient improves.

The patient thinks "My doctor must be very smart". Meantime, the girl outside that room in the thought-experiment is thinking "This man must be a brilliant Confucian scholar."

Part of being an intern is adjusting to all of this, losing some of your delusions of heroism, getting used to the fact that you're not going to be Dr. House, that you are at best going to be a very well-functioning gear in a vast machine that does often tedious but always valuable work.

Well, other people are. *I* plan to go into outpatient.

Starting tomorrow, I abandon this exciting world of urinary tract infections and broken legs and go into psychiatry full time. I'm looking forward to it, especially since psychiatry is a little slower-paced and more focused. But this year was meant to teach me some appreciation for the wider world of medicine.

And boy have I got it.

[Good luck to SSC commenters Athrelon and Laura and everyone else starting an internship or residency tomorrow, and congratulations to everyone finishing one up]

Searching for One-Sided Tradeoffs

Suppose you are an admissions official for a moderately prestigious college, which is neither the best nor the worst in your state. Your job is to look over people's SAT scores, high school GPA, and essays on How I Overcame Adversity, and then decide whether or not to admit them to your college.

And suppose that you have a team of subordinates who make the really easy decisions for you. Auto-reject the losers who show up drunk to their interview and spell your institution's name as "collej" on their applications, pass the rest on to you.

Your job probably doesn't matter. Yes, there will be some very high quality candidates – the kids with straight As, perfect SATs, and stories about how they personally stopped the civil war in Lebanon despite being born without legs. But they will be using you as their safety school, and whether you accept them or not they will be going to Harvard and you will never see them. You will only be deciding among a small band of students – those too smart to get auto-rejected by your subordinates, but not smart enough to go to a school better than yours.

Given that kids who are good at everything and kids who are bad at everything are equally unlikely to be your target population, your job reduces to choosing what tradeoffs to take. Do you want kids with great SAT scores but terrible grades, kids with great grades but terrible SATs, or kids with mediocre grades and test scores alike? How about kids with terrible grades and terrible SATs, but they're really really attractive and good at sports?

Even here your job won't matter *too* much. Your counterparts at Harvard will presumably be smart people who have a pretty good idea of how important test scores and grades are in terms of the Intangible Qualities That Make You Good At College. If a new study comes out showing that SAT scores determine your future but grades are meaningless, that study will make you want to shift to a high-SAT-low-grade model, but it will equally increase the high-SAT-low-grade kids' ability to get into Harvard, meaning that you will, to use an economics metaphor, have to buy SAT scores with grades at a lower exchange rate.

So basically no matter how competent you are as an admissions official, all of the kids entering your college will be about equally "good".

There is a fun legend I heard in a stats class – I don't know if it's true – of a psychology professor who got very excited about her new theory that the brain traded off verbal and mathematical intelligence – being better at one made you worse at the other. She got SAT Math and SAT Verbal scores from her students and found it supported her theory. A friend of hers did a replication at his college and found support for the the theory there as well.

But larger scale testing disconfirmed the theory. What the professors working off college samples were finding was that all of the kids in their college were equally "good", in a general sense, so excellence in any quality implied a tradeoff in other qualities. Suppose the professor worked at a mid-tier college – students with SATs much less than 1200 couldn't get in; students with SATs much more than 1200 could and did go to better schools instead. Then all her students would have SATs around 1200. Which meant a student with an SAT Verbal of 700 would have an SAT Math of 500, a student with an SAT

Math of 800 would have an SAT Verbal of 400, and boom, there's your "trade-off of verbal and mathematical intelligence". Obviously the tradeoff wouldn't be perfect, since there's random noise and since students are also trading off less obvious qualities like attractiveness, wealth, social skills, athleticism, musical talent, and diligence. But it would be more than enough for her to find her correlation if she was looking for it.

This suggests some odd strategies if we're looking for particular college students. If we want to find the dumbest students in a particular college, we might look at the football star – not because football stars are naturally dumb, but because plausibly a student who couldn't get in on his wits alone might make it in on the promise of helping the college team. If we want to find the smartest student in a particular college, we might look for someone on a scholarship – because perhaps she would otherwise be at Harvard, but was made less attractive to the Ivy League by her inability to pay them any money.

It also implies some weird strategies for admission officers. How do you maximize student quality when in theory all your job allows you to do is make tradeoffs between different subcharacteristics among students of the same quality? Aside from just hoping the occasional Harvard-caliber student accidentally stumbles into your office, I suggest three potential techniques: insider trading, bias compensation, and comparative advantage.

Insider trading is where you're just plain smarter than everyone else. Maybe you're a brilliant psychologist who has invented a test that invariably reveals students' true potential. You can find kids with terrible grades and terrible SAT scores

who will nevertheless shine. If you happen to luck into this position, you've got it made.

Bias-compensation is where you try to see if other colleges have biases that you can exploit. Sometimes this is simple and profitable. If Harvard is controlled by anti-Semites and auto-rejects all Jews, then you have a free shot to get Jews with 800 SAT Math, 800 SAT Verbal, *and* amazing football talent (though good luck finding Jews with amazing football talent). Once again, if you happen to luck into your competitors being stupid, you've got it made.

Sometimes it's not that easy, and you have to kind of spin someone else's preferences as "bias" when they might secretly have some wisdom behind them. For example, it is no doubt true that college admissions officials are influenced by student charm and social skills. So if you want, you can probably get smarter students if you go for the really really unpleasant students whom everyone dislikes as soon as they open their mouths. You can then declare "success" when your college gets a disproportionate number of academic awards, but unless you are a remarkably single-minded academic-award-maximizer, you may find that your college is kind of horrible now and other schools had pretty good reasons for rejecting these people.

Comparative advantage is where you decide you are going to have radically different priorities than anybody else. Maybe you want to be The Math School and become known for the quality of your math geniuses. So you nab all the students with 800 SAT Math and 400 SAT Verbal and then advertise the heck out of your students' mathematical acumen. There's also another sort of comparative advantage, where if you have a great sign language interpretation program and Harvard doesn't, you can advertise to deaf kids who maybe Harvard

doesn't want because they can't develop their talents effectively.

So let's generalize from college to the sorts of choices that we actually face.

In one of the classics of the Less Wrong Sequences, Eliezer argues that policy debates should not appear one-sided. College students are pre-selected for "if they were worse they couldn't get in, if they were better they'd get in somewhere else." Political debates are pre-selected for "if it were a stupider idea no one would support it, if it were a better idea everyone would unanimously agree to do it." We never debate legalizing murder, and we never debate banning glasses. The things we debate are pre-selected to be in a certain range of policy quality.

(to give three examples: no one debates banning sunglasses, that is obviously stupid. No one debates banning murder, that is so obviously a good idea that it encounters no objections. People *do* debate raising the minimum wage, because it has some plausible advantages and some plausible disadvantages. We might be able to squeeze one or two extra utils out of getting the minimum-wage question exactly right, but it's unlikely to matter terribly much.)

So there's some argument to be made that, like the admissions officer, our decisions aren't too important. I don't think things are quite that depressing. But, like the admissions officer, we will have to be clever if we want to figure out how to escape the seemingly iron law of tradeoffs.

I recently heard a Catholic guy condemn the "culture of death", which by his telling consisted of abortion, stem cells, euthanasia, and capital punishment. I'm in favor of three of those things, and I avoid a perfect four-out-of-four only on a

technicality: I can't support capital punishment until it gets better at sparing the innocent and maybe becomes more cost-effective.

My near-unaninimous support for culture-of-death issues seems unlikely to be a coincidence, and indeed it isn't. I have a deep philosophical disagreement with the Catholics here – they think life is a terminal value, I think life is only valuable insofar as it gives certain goods associated with living.

This means from my point of view, the Catholics have a bias in their trade-off arithmetic. They are the equivalent of the anti-Semitic Harvard leadership, who have given me this great gift of trade-off-free students. Just as learning the Harvard leadership is anti-Semitic makes me suddenly want to accept all Jews as a tradeoff-free utility gain, learning that a large portion of the electorate is biased against death means that certain death-related policies can be tradeoff-free utility gains to me.

I will add one more political example. I've previously proposed sticking lithium in the water supply as an intervention to promote psychiatric health. People are super creeped out by this – and in fact, so am I, a little bit. But this is encouraging! If people's response was "actually, we have proof that these quantities of lithium hurt cardiac health" we'd be faced with a useless tradeoff – X psychiatric health against Y cardiac health – and so a policy we'd be squeezing a couple measly utils out of depending on which way the tradeoff went. But if their response is "I see no particular downside, but I am very creeped out by it", then this is like learning Harvard is anti-Semitic – an explanation for why other people haven't gobbled up a possible advantage, and a neon sign pointing out potential tradeoff-free gains for you.

We can also use this framework to evaluate life hacks.

Life hacks are touted as “little-known techniques you can use to improve your life”. There are two ways something can fail to be a life hack – either it becomes universally known, or it fails to improve anything. These form a pre-selection kinda like a college selecting students of a certain quality, or a country debating issues of a certain quality. If an intervention was obviously great, then either you’d already do it (think “sleeping at night” or “working at a job to earn money”) – or you would at least feel guilty for not doing so (think “diet and exercise”). If an intervention was useless, no one would call it a life hack (think “hitting yourself on the head with a baseball bat every day”). Life hacks are the things that are sort of in between, where there seem to be some benefits, and also some costs in terms of time and energy and money, and you’re not sure if they’re worth looking into or not.

If you want to do better than trade off your time and energy for the occasional small benefit, you need a theory of why that might be possible.

Every life hacker wants to be an insider trader – someone who is able to outperform competitors with more resources by being a little savvier about biology and psychology. And probably some are. But unless you are the first scientist to discover a new supplement, or the first psychologist to discover a new technique, your trades aren’t that insider and you’ll eventually have to explain why no one else has adopted them.

And most life hackers pay lip service to comparative advantage: “Everyone has their own individual biology and their own set of problems, what works for you may not work for everyone else.” This is pretty plausible. It suggests the

reason the whole world isn't adopting life hacks is because there's a very high startup cost, where you've got to sort through a hundred different things and find the ones that work for you and the ones that don't, and nobody can do this for you, and if you're not very smart you'll get it wrong.

Another form of comparative advantage is willpower. Maybe no one else is doing weight lifting because they don't have the determination to go to the gym three times a week. This is a fine theory – plausible even – but it's interesting to see how many of the people who confidently assert their own comparative advantage then buy a gym membership but end up not having the determination to go three times a week.

But in terms of using a tradeoff-based framework to help inform the decision of what lifehacks to try, it seems most promising to consider opportunities for bias compensation.

Like insider trading, bias compensation is claimed a lot more often than I think it can be supported. The polyphasic sleep crowd, for example, tell you that you can increase your free time per day – and all you need to do is stick to a very strict schedule, be very tired for a long time while you're working out kinks, and abandon all hope of a social life or flexible schedule. To me this seems a lot like the admission official with the bright idea of admitting unpleasant low-social-skills kids: it sounds good if you're only thinking about the most easily quantifiable results, but when you actually try it you tend to regret it very quickly.

Can we find anything more promising? I think that people are unnecessarily pessimistic about nootropics because they are scared of taking drugs. Fear of taking drugs is an excellent and rational fear to have, but if you happen to lack it that fear, or you have enough comparative advantage in pharmacological

knowledge / research ability that you can justifiably be less afraid of taking drugs than everyone else, then this starts to look like the lithium-water example: getting free utility by abandoning your sense of creeped-out-ness.

But if you're going to force me to give you an example of something I actually did differently because of thinking about tradeoffs, I'll have to go with "try bacopa".

Bacopa is a memory-enhancing drug that performs very well in studies. But it's rarely used and it only got a middling ranking on my survey. I think this has something to do with having to take it for three months before it has any effect. Talk about trivial inconvenience. Most people don't want to bother, so it remains largely uninvestigated, and the nonsuperabundance of bacopa use stands explained without resorting to it being a bad drug or having other tradeoffs we really don't want. So using it – if you can stand the three month waiting period – has a higher-than-otherwise-expected likelihood of being free utility.

Me? I tried to start taking bacopa, but it gave me terrible diarrhea and I had to stop. Another tradeoff! That should just increase its expected psychological benefits!

Last, something on the lighter side: an article going around the Internet recently claims houses on streets with mildly rude names (example: "Slag Lane") apparently cost £84,000 less than control houses on more properly named streets (the article does not give me enough information to rule out hypotheses like poor people being more willing to give their streets rude names). If you don't care about what your street name is called, this might be another potential free trade-off – buy a house on Slag Lane and save $100,000+. Or buy a house that's supposed to be haunted if you don't believe in ghosts. Or buy

a house near a prison with a very low escape rate because you trust the statistics and other people don’t.

Do Life Hacks Ever Reach Fixation?

In Searching For One-Sided Tradeoffs, I argued that people's "life hacks" probably occupy a restricted range. If the life hack had *nothing* going for it, it would never become popular and you wouldn't hear about it. If the hack had *everything* going for it, you would have heard *more* about it. If there were something that really doubled energy levels and increased IQ and cured shyness and made you lose ten pounds, the front page of the New York Times would be "Man Discovers Amazing Life Hack", and it would be all over the medical journals and the talk shows and so on. It wouldn't have to be pushed by some guy with a blog who says it "changed his life".

…except that then I tried looking for examples of such and came up blank. The example I ended up giving, "sleeping at night", was a biological imperative that was never really "discovered", per se.

Compare to genetics. If there's a mutation that gives even a small benefit, it predictably reaches fixation in the population (where every single organism has it) after a certain number of generations.

Compare also to other kinds of ideas, like technology. When a new technology (let's say the cell phone) is invented, it starts with a group of early adopters. As the technology gradually gets better and cheaper, and people notice that cell phone users have a big advantage over non-users, new people buy cell phones. Eventually it reaches the point where the cell-phone-less are at a big disadvantage, and even the grumpy old holdouts like myself are forced to purchase them. Even if we never reach *literal* fixation because of the Amish, the indigent,

etc, there's still a point in which having a cell phone seems to become the default state.

The same is true in the economy. One business gets a bright idea, like outsourcing to China or something, they get rich and outcompete their rivals, their rivals pick up on the idea, and eventually businesses-that-don't-outsource-to-China gets reduced to a weird niche market.

This should be able to work with lifehacks. Whether it's students trying to get the best grade, workers trying to be most productive, or suitors trying to appear most attractive, people compete with each other *all the time*. If there were some meme that consistently offered its users an advantage in productivity or energy or even mood, it ought to reach fixation as surely as new technologies or business practices.

And I can't think of any that have.

Some possible explanations:

1. There are no exceptionally good life hacks. The human body and brain are optimized really really well, or else have really really strong tendencies to return to equilibrium after a disruption.

2. Life hacks, as a category, have some characteristic that makes fixation an unreasonable goal for them. Maybe there is so much variation in people that no lifehack can ever improve more than a small percent of them. This seems like a less bleak version of (1) – the stuff everyone has in common is optimized really really well, but there are some individualized flaws you can pick off on a person-by-person basis.

3. Life hacks as a category didn't exist until kind of recently, or it if did they weren't as good as modern life hacks. Even

though there are some great ones out there now, they haven't existed long enough to achieve fixation.

4. All the genuinely useful life hacks take work, and people are really bad at doing work, so nothing that takes work can ever achieve fixation. The level of work it takes to understand a cell phone or computer doesn't count; these life hacks take more work, or different kinds of work.

5. Some life hacks have totally reached fixation and I'm just too stupid to think of them. Or – life hacks that reach fixation become so entrenched that it's very hard to think of them as lifehacks any more. Compare the genetics student who says "No mutations have ever reached fixation in the human population, and I know this because most of the people I see aren't mutants."

The last explanation seems most promising, which means I should probably look harder for fixated life hacks.

There are some things I want to exclude right away. New technologies like the cell phone can reach fixation, but I don't think I'd want to call them life hacks; I'd rather limit the term to non-medical interventions or at least technologies specifically related to health and productivity. Certain ideas like religion have reached fixation in their populations, and it would be fascinating to think of in what senses those are life hacks, but I don't think that's where we're going here. I'm looking specifically at things that act directly to raise energy levels, intelligence, social skills, or organizational ability.

I will grudgingly accept three-ring binders, to-do lists, calendars, and filing cabinets as *sort of* examples – even though I don't use a calendar or to-do list and it doesn't seem to have left me unable to compete with the rest of humanity,

and even though these all fall into a sort of general “keep organized by writing things down and sorting them” category.

I will grudgingly accept backpacks, briefcases, and the like, even though “things that hold other things” seems to be a pretty basic human invention and if we have to go back to the Paleolithic before getting a genuinely useful life hack we are doing very poorly indeed. This might also be a piece of technology which escapes that category only through the cheap trick of going so far back that it doesn’t seem like a technology anymore.

I will grudgingly accept “diet and exercise”, since even people who are bad at diet and exercise probably eat better and exercise more than they would if they were unaware that diet and exercise were things they should do. But I don’t know if this was ever really “discovered” or if it got a lot of help from a biological imperative.

I will grudgingly accept “take a deep breath and count to ten in order to not get angry”, since everyone seems to know about it.

But none of these seem to fall into classical life hack categories like “thing that a man with slick hair teaches a class on, telling you that it will change your life”, or “thing that you can buy at the Sharper Image”. And they all seem pretty old. Cell phones took like fifteen years to achieve fixation; how come for life hacks we have to look all the way back to whichever caveman first realized you could carry tools in a sack made of animal skin?

EDIT: @mjdominus on Twitter proposes caffeine. That sounds right to me.

Polyamory is Boring

I.

I remember explaining polyamory to my father when I met him in Utah. He just shrugged and said "I guess I'm too old-fashioned for that sort of thing to make sense."

I feel blessed to have a father with the rare skill of being able to generate "I am old-fashioned" as a counter-hypothesis to "other people are evil". But more than that, I sympathize with his response. I sympathize with it because it was exactly my response when Alicorn told *me* about polyamory two years ago or so (I can't remember if I got it from IM conversations with her or from reading her essay). For a twenty-eight year old, I am *really* good at sighing and saying "Kids these days!" in a despairing tone, and that was about my response to the whole polyamory concept.

And now seven months after moving to Berkeley I'm dating three people.

II.

What changed? It just started seeming *normal*. I was going to make an analogy to desegregation here, how white people thought having black kids in their schools would be a disaster, and then it happened, and the world didn't collapse into a hell dimension or anything, and after a few years it just seemed like the normal order. But that metaphor is too weak: there's still racism, a black kid in an all white school district probably feels really out of place, there are still even fights over segregated proms.

So better analogy. Imagine a space-time rift brings a 19th-century Know-Nothing to your doorstep. He starts debating

you on the relative merits and costs of allowing Irish people to mix with the rest of American society. And you have a hard time even getting the energy to debate him. You're like "Yeah, there are some Irish people around. I think my boss might be half-Irish or something, although I'm not sure. So what?" And he just sputters "But…but…Irish people! It's not right for Irish and non-Irish people to mix! Everyone knows that!" And not only do *you* not think that Irish people are a Big Deal, but you're about 99% sure that after the Know-Nothing spends a couple of months in 21st-century America *he's* going forget about the whole Irish thing too. There's just no way someone seeing how boring and ordinary Irish-Americans are could continue to consider worrying about it a remotely good use of their time.

In fact, this Know-Nothing would have two strikes against him if he tried to hold onto his philosophy. First, there's the empirical strike. Whatever his predictions of doom – Irish immigration would impoverish the country, Irish immigration would lead to the US being annexed by the Vatican – those predictions have clearly been disconfirmed. Second, there's the psychological strike. He would be exposed to so many perfectly normal Irish people that his brain would have trouble even maintaining them as a separate category. It's like the difference between your association for Chinese people being Fu Manchu versus your association being your neighbor John Chang who speaks perfect English and has a job at Google.

This was my experience with poly people upon moving to Berkeley. Alicorn makes a big deal about poly-hacking and having to valiantly overcome some sort of strong natural tendency to switch from monogamous to polyamorous relationships. This wasn't really my experience at all. It just seemed like once the entire culture was no longer uniting to

tell me polyamory was something bizarre and different and special, it wasn't. And then it started to look like a slightly better idea to take part in it than to not take part in it. So I did.

III.

I didn't even remember how weird it seemed to everyone else until the last few weeks. First I had to explain it to my father. Then someone commented on a blog post of mine with something about polyamory, spelling it poly-"amor"-y the whole time, as if there couldn't possibly be any real love involved.

The plural of anecdote is not "data". But the singular of anecdote is "enough data to disprove a universal negative claim". So I will just say that Alicorn and Mike are probably the best couple I have ever seen. I have lived with them for seven months now and never once have I seen them get in a fight (I know there is way more to being a couple than not fighting but I'm trying to think of objective numerical evidence I can report here beyond "if you know them, you know what I mean"). They both seem to love and appreciate each other just as much if not more as they did when I first met them. They both go *way* out of their way to make the other happy, and although part of this is just that they're both very nice people who go out of their way to make *everybody* happy, I think there's got to be some love involved there too. They are engaged, working on the "getting married" thing, and have every intention of having lots of children and staying together for at least one lifetime.

And all this despite Mike having two other girlfriends and Alicorn having three other boyfriends including one who lives with her. I can't even get angry with people who say polyamory is incompatible with true love. They're just

empirically wrong, like someone who remarks confidently that hippos have six legs. They're not evil or even deluded. They just obviously haven't seen any hippos. You don't really want to argue with them so much as take them to a zoo, after which you are confident they will realize their mistake.

IV.

The other thing people always bring up is the jealousy issue. I feel like the correct, responsible thing to say at this point would be "Yes, of course everyone experiences jealousy, and it's hard for the first few months or years, but eventually you just learn to live with it and the sacrifice is worth it."

But the responsible answer is wrong, and the incredulous-stare answer is right. At least in my very limited experience, jealousy is a paper tiger, sort of the post-9/11 al-Qaeda of emotional states. You spend all this time worrying about it and preparing for it and thinking it is going to be this dreadfully imposing enemy, and in the end it sends one guy with a bomb in his shoes onto a plane, whom you arrest without incident.

I know this hasn't been anywhere close to the experience of all polyamorous people, but it's my experience and that of the people I've talked to most about this.

My roommate Mike dates the same three people I am dating, including Alicorn who also lives with us (this is not normal for polyamory, and all three people started dating Mike and then met me and started dating me too, so I guess the moral of the story is to think very hard before accepting me as a roommate). I cannot think of a single problem I have ever had with Mike, which I guess is also sort of incredulous-stare and which exceeds my normal standards for *roommates* let alone roommates-whose-three-girlfriends-I-am-dating. None of those three people have had any noticeable-from-the-outside

jealousy about any of the others. Two weeks ago, Mike and I took all three of our mutual girlfriends on a group date to Sausalito. It went really well, everyone got along, and it is something we would do more often if not for scheduling and travel issues (also, Sausalito is really expensive).

I *once* felt a small pang of jealousy when one of my girlfriends was having a very public display of affection with a non-Mike person I didn't know quite so well. But I get upset with/jealous of public displays of affection in general, even among people I don't know, and it's very hard for me to disentangle this feeling from jealousy and it could have just been my imagination.

As opposed to this tiny-to-nonexistant role of jealousy, I think pretty much everyone here has experienced compersion. Compersion is the opposite of jealousy, being really happy for your partner when they meet someone new and they are obviously happy. Mike and Alicorn are really good at compersion (Mike helped set me up with his girlfriend Kenzi and was really glad it worked out) and some of this has rubbed off on me. It is a good feeling and it makes you feel good to have it. If there is a Heaven, I assume compersion will be a big part of its emotional repertoire.

V.

I don't drink much, not because I'm especially virtuous but because I hate the taste of alcohol and the atmosphere of bars and parties. In the same way, I'm not promiscuous, not because I'm especially virtuous but because I'm sort of borderline asexual. I like cuddling people, kissing people, falling in love with people, petting people's hair, writing sonnets about people, and a few things less blogaboutable, but having sex isn't an especially interesting experience for me. I

treat it kind of like watching a chick flick – something one might do to get the nice warm feeling of doing romantic things and bonding as a couple, but wait a second why the heck is she kissing him now and that scene made no sense and THIS MOVIE HAS NO PLOT HOW DID IT MAKE $100 MILLION AT THE BOX OFFICE?

And I'm sorry for subjecting random people to details of my sex life, but I'm trying to establish credibility here for what I want to say next. What I want to say next involves the perception – I had it and a lot of other people seem to have it – that polyamory is about having sex with lots of people and monogamy is about having close loving relationships. And once again this is not my experience at all.

If you just want to have lots of sex instead of having a loving relationship, there are many ways to do it that are much more socially acceptable than polyamory. You can be one of those bachelors who "plays the field" and "doesn't get tied down". You can be in an "open relationship" or be "swingers". All of these are way easier than polyamory; if your goal is sex, they're also more effective.

Polyamory is almost the opposite of this. It's for people who *aren't* just into sex, for people who realize they could get sex without relationships with a lot less deviation from social norms but are really into the relationship part of things.

Here I will say maybe the only note of personal uncertainty or concern you're likely to get in this essay, which is that I don't know whether I could have maximally-close relationships with multiple people simultaneously. That is, I don't know if I could date three people and love all of them as much as my parents love each other, or other social models for very good relationships (the Obamas? Now I'm foundering on who our

non-fictional archetypes for very good relationships are) love each other. I'm not sure whether this would satisfy some deep human need for what you might politically-incorrectly call "mutual ownership". And I'm definitely not sure (though I think it's likely, certainly more likely than the skeptics would) that this is a great structure for child-rearing.

In practice none of this matters, because driven by some innate urge most polyamorous people I know end up having one "primary" relationship along with whatever others they are involved with. Mike and Alicorn are each other's primaries, and that is going to develop into being each other's spouses, and what I said above about them definitely having achieved that level of maximum-closeness remains true. This form of polyamory seems to me to be "monogamy plus", keeping all of the advantages of monogamous relationships and ending out strictly superior. Sometimes this develops into people being so into each other that they just aren't interested in other relationships because it takes away time they could be spending with their primary partner, but I haven't noticed any differences in the quality of relationships where this happens and ones where it doesn't.

I have heard of polyamorous communities where this is not how things are done, where people don't have primaries, where they are just this complicated mass of partners without anything that looks like a traditional relationship. I predict I would not like this; something in me recoils from this situation. But that could just be more prejudice that would look as dumb as a Know-Nothing in the 21st century once I saw it up close. I'm pretty willing to take the Biblical tack on this one: "He who is able to accept it, let him accept it". But I'm pretty sure I'm not of that number.

Can You Condition Yourself?

A friend recently told me about a self-help tactic that has become popular in the circles I move in: the idea of applying behaviorism to yourself (sometimes called "training your inner pigeon"). The idea is you give yourself rewards when you do things you want to do more of, and your brain works its magic and reinforces the activity.

When I first heard about this, my thought was "No way that is ever going to work". I have always been under the impression that conditioning is kind of like tickling. You can't tickle yourself. You'd be *expecting* it.

Let's start by distinguishing a couple of possibilities:

1) This process doesn't work at all

2) This process works by making you want the reward. Suppose you promise yourself a candy bar each time you do homework. You are hungry and want the candy bar, but you would feel bad if you ate it without doing homework. Therefore, you grudgingly do homework to get at the candy bar.

3) This process works by changing your urges and desires. After eating a candy bar each time you do homework, your brain associates homework with a nice, delicious-feeling, and you enjoy doing homework more from now on.

Let's start with 3, the most encouraging possibility. This gains a little support from the Little Albert experiment. Here, a baby who had no particular fear of rats was exposed several times to rats plus loud, terrifying noises. Eventually the baby came to fear rats, even without the noise, presumably because the fear of the noise had generalized onto the rat through association.

It's easy to see how this could mean something like the happiness of candy-bar-eating generalizing to homework. Nevertheless, I believe this argument proves too much.

Every evening, I sit down at the table, get a plate and some silverware, and eat dinner. It's usually something I really like, and it usually includes dessert, which I like even more. If eating good food isn't rewarding, I don't know what is, and sure enough I rarely skip dinnertime.

However, if for some reason I don't have dinner – maybe I've promised my friends I'll go out with a late dinner for them and so I can't stuff myself first – I do not feel the *slightest* urge to sit down at the dinner table with a plate and sort of move my silverware around in the air making little eating motions, and when I tried it (empiricism!) I did not find it at all pleasant.

Take a second to think about how weird that is (the result, not me trying the experiment). Sitting at the table and moving my silverware, in conditions exactly like these, has been quickly associated with reward every single time I've done it in the past, for decades, ever since I learned to feed myself. But I don't feel even a *little* bit of urge to do this. None at *all*. You may generate additional examples at your leisure, but the point is that just being consistently associated with a positive reinforcer in a low time-delay way does not make a neutral activity (let alone an actively unpleasant activity) become desirable.

What happened with Little Albert, then? First of all, he was classical conditioning and not operant conditioning. Second of all, Albert had no understanding or control over what was going on. Each time he heard the noise, he was very surprised – he was receiving a new fact from the Universe. But it wasn't information he understood; he had no idea what the connection

between the rat and the noise was and whether it would recur. He just knew that there was some mysterious rat -> noise connection.

Compare this to me eating dinner. The connection between sitting down and eating dinner is not at all a new fact fed me by the Universe; it's something I plan myself. And it is not mysterious whether any given sitting and silverware-waving will reward me; I know it will reward me if and only if I am planning to eat dinner. Therefore the brain does not think of silverware-waving as an activity that might, who knows, lead to reward in the future.

(one might object that my inner pigeon – or lizard brain, to mix animal metaphors – doesn't share my complex explicit knowledge of the reward structure of dinner-eating. But the little I know of the brain's reinforcement mechanism suggests that reinforcement learning is based on *surprise* – technically the difference between predicted and observed values of some complicated Bayesian equation encoded in dopaminergic neurons or something – and that this system is actually quite good at predicting expected reward from an action, within certain limits)

So (3), the hypothesis that the reward will cause me to start enjoying homework, seems wrong. What about (2) – "I don't like homework much, but at least I get some candy out of it"?

Here there's a ceiling on how much the candy can reinforce your homework-doing behavior, and that ceiling is how much you like candy.

Suppose you have a big box of candy in the fridge. If you haven't eaten it all already, that suggests your desire for candy isn't even enough to reinforce *the action of going to the fridge, getting a candy bar, and eating it*, let alone the much more

complicated task of doing homework. Yes, maybe there are good reasons why you don't eat the candy – for example, you're afraid of getting fat. But these issues don't go away when you use the candy as a reward for homework completion. However little you want the candy bar you were barely even willing to take out of the fridge, that's how much it's motivating your homework.

Maybe you say "I will allow myself exactly one candy bar a day, but only if I finish my homework". Even if you can stick to this rule, here the candy bar becomes an *extrinsic reward* motivating the homework. We all know what happens with extrinsic rewards – overjustification effect! You gradually start interpreting the task at hand as an annoying impediment to getting the reward, lose your intrinsic motivation, and as soon as the reward is removed, you're even less willing to do the task than before.

So both (2) and (3) are pretty unlikely. That leaves us with (1) – don't even bother.

Luckily, my friend helpfully clarified that this wasn't what her class taught at all (I think maybe they originally tried this, but considerations like the ones I mentioned convinced them to change?). Their new policy is that you should reinforce yourself with a "victory gesture" – for example, pumping your fist and shouting "YEAH!" and visualizing an image corresponding to your success and trying to feel really good about yourself.

So for example, as soon as you sit down to start your homework, you make the victory gesture and imagine yourself graduating *summa cum laude* from school, and then you feel really good and have reinforced the behavior of sitting down

to do your homework. And maybe you do it again when you finish, because peak end rule.

She claims a few benefits of this method. First, it's *very fast*, so you can reinforce things right as they happen instead of with time delay which gives your brain enough time to lose the connection. Second, it's intrinsic, so it's not going to sap your natural motivation the same way the candy bar might.

I understand the claim that rewards delivered *very immediately* after a stimulus can work better for conditioning – I was referred to a couple of papers proving this, though I don't remember them. But I notice I am confused. When we have good examples of *real* conditioning, immediate reward isn't especially important. For example, people often use the language of behaviorism to talk about addiction, say alcoholism. But the chemical rewards of getting drunk don't manifest until a little while after you've had your first beer – certainly not within a split second – and certainly alcoholism can reinforce even longer term behaviors, like leaving home and going to the bar. Pornography is another good example of effective behaviorism, but going to a porn site gives only delayed rewards – first you have to find a video you like, then you have to wait for it to buffer, then you have to sit through the boring part where the nice lady and the plumber are discussing the best ways to fix her faulty pipes, and so on. It seems that when we have a real effect that definitely works, immediacy is *not* required (indeed, if it were humans would have a lot of trouble learning anything but the most basic reflexes).

But okay. Ignore that. It would really really really really bad mind design to allow your own consciously generate-able emotions to feed back into the reinforcement mechanism.

Start with one obvious point. I said the candy bar couldn't be much of a reinforcer if you otherwise left it in the jar without eating it. The same seems broadly true of a victory gesture. I don't feel the *slightest* urge to perform a victory gesture, and having tried it empirically I don't feel the slightest urge to repeat it. This bodes poorly for its ability to be a strong reinforcer.

And over several billion years of evolution, the brain has every incentive to get rid of that behavior if indeed it was ever possible. Imagine a world in which our own thoughts and feelings can be strongly reinforcing. You're a caveman, encountering a saber-toothed tiger. You have two choices. You can either feel fear, which is an unpleasant emotion. Or you can feel happiness, which is a pleasant emotion. First you try feeling fear, but that's unpleasant! You don't like fear! The feeling of fear is negatively reinforced and your brain learns to stop feeling it. Then you try happiness! You like happiness! The decision to feel happiness is positively reinforced. Yes, you decide, saber-toothed tigers are wonderful things and you are overjoyed there is one in front of you getting into a pouncing position and licking its lips and…well, this caveman isn't going to live very long.

From the little I know about the reward system, it seems to operate on a basis of predicting pleasure level, then upregulating actions that result in world-states that seem more pleasurable than predicted and downregulating actions that result in world-states that seem less pleasurable than predicted. I don't think you can prevent the "I'm going to do my victory gesture!" part of you and the "I'm going to predict my pleasure at time t+1" part of you from talking to each other, I don't think internal pleasure is as reinforcing as external

world-state results, and I don't think the pleasure of making a victory gesture is strong enough to do much anyway.

…there were a lot of "I thinks" in that paragraph. Do we have any evidence here?

The literature on this is hiding under the obscure term "self-consequation", and unfortunately it is all from Scientific Prehistory, ie the 1970s and 1980s before journal articles were uploaded to the Internet. I am able to find this full study, which does pretty much exactly the experiment listed at the beginning of this post – feed people candy in return for studying – and finds that it helps only if other people are there keeping them honest. But I am also able to find this abstract, which appears to be from a study showing the opposite – some kind of benefit – but is totally unavailable on the Internet. Both studies seem to refer to a long literature supporting their result and (sigh) neither seems aware of the other's existence. However, I am more skeptical of the second, both because I can't see it and because I worry that experimental protocols aren't *real* self-reinforcement. That is, if an experimenter gives you *their* bag of candy and tells you to reinforce yourself by eating some when you do something good, that's still different from using your own bag of candy and coming up with the idea on your own, even if the experimenter is out of the room when you're working.

I will still try the technique, because it seems low cost and potentially high value. Really high value, actually. So high value that I would have expected the first person to get it right to take over the world. This is turning into another argument against it, isn't it?

But yeah, as I was saying, I still intend to try the technique, even though it won't be a very well-controlled experiment.

And I’m glad I heard the idea for reminding me how little I know about behaviorism.

Wirehead Gods on Lotus Thrones

One vision of the far future is a "wirehead society". Our posthuman descendants achieve technological omnipotence, making every activity so easy as to be boring and meaningless.

The pursuit of material goods becomes a waste. A nanofactory or a quick edit to a virtual world can already give you a mansion the size of a planet. Although economic activity may still exist in competition for computing resources, all beings in these competitions will be smart enough to behave perfectly optimally (and therefore in a way that makes even the illusion of free will impossible) and so first-mover advantages will be insurmountable. Economic differences will compound on the sub-second scale until different classes are so far apart that competition becomes impossible.

When sports risk becoming contests of who can enter the higher integer in the $athletic_ability variable of the computer that determines the universe, the World Anti-Doping Agency says everyone must compete using their original human bodies – assuming such things even exist at that point. But neither spectators nor athletes care about the result, since everyone is smart enough to simulate the game in their minds on a molecule-by-molecule basis long before it happens and determine the outcome with perfect accuracy – turning the actual competition into a meaningless formality.

Works of art become gradually less interesting; everyone can extrapolate back from the appearance of a painting to exactly what the mental state of the artist must have been at the time it was painted. Nor does the art enlighten, since the conceptual organization of everyone's mind is already optimal and the

only intellectual differences between entities are insurmoutable ones of available computing resources.

As for Science, everything was discovered long ago. If it hasn't been, discovering it is a brute-force application of the best-known Bayesian reasoning algorithms.

(And developing better algorithms is also a brute-force application of the best-known algorithm-discovery algorithms.)

Even in the most utopian such world – one where the dominant minds are concerned with maximizing the happiness of everyone else – it sounds pretty boring.

One approach is the imposition of artificial limits. Entities can deliberately refuse to use their full cognitive capacity and so experience uncertainty, choice, and feelings other than that of algorithmically choosing purpose-appropriate algorithms. Maybe some entities will deliberately take on human brains and bodies, and interact with other such entities in a human-level world in order to operate at the level with which their value system is most comfortable. Maybe in order to avoid the temptation to call on their full omnipotence every time they experience a little pain or hardship, they will deliberately "forget" their posthuman status, living regular human lives utterly convinced that they are in fact regular humans.

(I assign moderate probability that this has already happened)

Other entities may have no time for such games. They may cope with the ennui of posthuman existence by reprogramming away their capacity for ennui, with the absence of aesthetic or scientific outlets by programming away their desires for such. Instead, they just reprogram their brains to be deliriously happy all the time no matter what, and spend their time sitting around enjoying this happiness.

The futurist community calls this “wireheading”, after an experiment in which rats had an electrode hooked up to the reward system of their brain which could be stimulated by pressing a lever. The rats frantically tried to stimulate the lever as much as possible in preference to doing anything else including eat or sleep (they eventually died). Stimulating the reward center directly was much more attractive than other activities which might result in some indirect neural reward only after work.

The same pattern occurred in humans, specifically chronic pain patients who had similar wiring installed in their heads in the hopes that it might alleviate their problem:

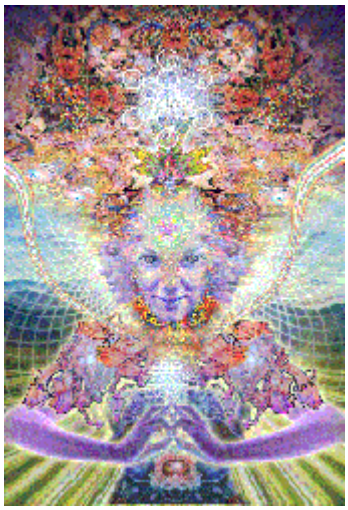
> At its most frequent, the patient self-stimulated throughout the day, neglecting personal hygiene and family commitments. A chronic ulceration developed at the tip of the finger used to adjust the amplitude dial and she frequently tampered with the device in an effort to increase the stimulation amplitude. At times, she implored her to limit her access to the stimulator, each time demanding its return after a short hiatus. During the past two years, compulsive use has become associated with frequent attacks of anxiety, depersonalization, periods of psychogenic polydipsia and virtually complete inactivity.

It’s unclear to what degree these wires are making the subject so stupendously happy that she desires to maintain her bliss, or whether they’re instilling compulsive behavior. Likely there are some elements of both – just as in wireheading’s more prosaic younger sister, everyday drug use. But drug use is messy, and wireheading is perfect.

Wireheading is commonly considered an ignoble end for the human race – our posthuman descendants reduced to sitting in dingy rooms, taking never-ending hits of some ultra-super-drug, all their knowledge and power lying fallow except the tiny fraction necessary to retain delivery of the ultra-drug and pump nutrients into their veins.

On the one hand, it probably beats desperately trying to figure out something to do more interesting than setting your $athletic_ability statistic to 3^^^3 and playing sports. On the other, it's hard not to feel contempt for beings that choose such a pathetic existence.

But I recently realized how unstable my contemptous feelings are. Imagine instead our posthuman descendants taking the form of Buddhas sitting on vast lotus thrones in a state of blissful tranquility. Their minds contain perfect awareness of everything that goes on in the Universe and the reasons why it happens, yet to each happening, from the fall of a sparrow to the self-immolation of a galaxy, they react only with acceptance and equanimity. Suffering and death long since having been optimized away, they have no moral obligation beyond sitting and reflecting on their own perfection, omnipotence, and omniscience – at which they feel boundless joy.



Pictured: ultimate reality

I am pretty okay with this future. This okayness surprises me, because the lotus-god future seems a lot like the wirehead future. All you do is replace the dingy room with a lotus throne, and change your metaphor for their no-doubt indescribably intense feelings from “drug-addled pleasure” to “cosmic bliss”. It seems more like a change in decoration than a change in substance. Should I worry that the valence of a future shifts from “heavily dystopian” to “heavily utopian” with a simple change in decoration?

Don’t Fear the Filter

There’s been a recent spate of popular interest in the Great Filter theory, but I think it all misses an important point brought up in Robin Hanson’s original 1998 paper on the subject.

The Great Filter, remember, is the horror-genre-adaptation of Fermi’s Paradox. All of our calculations say that, in the infinite vastness of time and space, intelligent aliens should be very common. But we don’t see any of them. We haven’t seen their colossal astro-engineering projects in the night sky. We haven’t heard their messages through SETI. And most important, we haven’t been visited or colonized by them.

This is very strange. Consider that if humankind makes it another thousand years, we’ll probably have started to colonize other star systems. Those star systems will colonize other star systems and so on until we start expanding at nearly the speed of light, colonizing literally everything in sight. After a hundred thousand years or so we’ll have settled a big chunk of the galaxy, assuming we haven’t killed ourselves first or encountered someone else already living there.

But there should be alien civilizations that are a *billion* years old. Anything that could conceivably be colonized, *they* should have gotten to back when trilobytes still seemed like superadvanced mutants. But here we are, perfectly nice solar system, lots of any type of resources you could desire, and they’ve never visited. Why not?

Well, the Great Filter. No knows *specifically* what the Great Filter is, but *generally* it’s “that thing that blocks planets from growing spacefaring civilizations”. The planet goes some of

the way towards a spacefaring civilization, and then stops. The most important thing to remember about the Great Filter is that it is *very good* at what it does. If even one planet in a billion light-year radius had passed through the Great Filter, we would expect to see its inhabitants everywhere. Since we don't, we know that whatever it is it's *very* thorough.

Various candidates have been proposed, including "it's really hard for life to come into existence", "it's really hard for complex cells to form", "it's really hard for animals to evolve intelligent", and "actually space is full of aliens but they are hiding their existence from us for some reason".

The articles I linked at the top, especially the first, will go through most of the possibilities. This essay isn't about proposing new ones. It's about saying why the old ones won't work.

The Great Filter is not garden-variety x-risk. A lot of people have seized upon the Great Filter to say that we're going to destroy ourselves through global warming or nuclear war or destroying the rainforests. This seems wrong to me. Even if human civilization does destroy itself due to global warming – which is a lot further than even very pessimistic environmentalists expect the problem to go – it seems clear we had a chance *not* to do that. A few politicians voting the other way, we could have passed the Kyoto Protocol. A *lot* of politicians voting the other way, and we could have come up with a really stable and long-lasting plan to put it off indefinitely. If the gas-powered car had never won out over electric vehicles back in the early 20th century, or nuclear-phobia hadn't sunk the plan to move away from polluting coal plants, then the problem might never have come up, or at least been much less. And we're pretty close to being able to colonize Mars right now; if our solar system had a slightly

bigger, slightly closer version of Mars, then we could restart human civilization anew there once we destroyed the Earth and maybe go a *little* easy on the carbon dioxide the next time around.

In other words, there's no way global warming kills 999,999,999 in every billion civilizations. Maybe it kills 100,000,000. Maybe it kills 900,000,000. But *occasionally* one manages to make it to space before frying their home planet. That means it can't be the Great Filter, or else we would have run into the aliens who passed their Kyoto Protocols.

And the same is true of nuclear war or destroying the rainforests.

Unfortunately, almost all the popular articles about the Great Filter miss this point and make their lead-in "DOES THIS SCIENTIFIC PHENOMENON PROVE HUMANITY IS DOOMED?" No. No it doesn't.

The Great Filter is not Unfriendly AI. Unlike global warming, it may be that we never really had a chance against Unfriendly AI. Even if we do everything right and give MIRI more money than they could ever want and get all of our smartest geniuses working on the problem, maybe the mathematical problems involved are insurmountable. Maybe the most pessimistic of MIRI's models is true, and AIs are very easy to accidentally bootstrap to unstoppable superintelligence and near-impossible to give a stable value system that makes them compatible with human life. So unlike global warming and nuclear war, this theory meshes well with the low probability of filter escape.

But as this article points out, Unfriendly AI would if anything be even more visible than normal aliens. The best-studied class of Unfriendly AIs are the ones whimsically called

“paperclip maximizers” which try to convert the entire universe to a certain state (in the example, paperclips). These would be easily detectable as a sphere of optimized territory expanding at some appreciable fraction of the speed of light. Given that Hubble hasn’t spotted a Paperclip Nebula (or been consumed by one) it looks like no one has created any of this sort of AI either. And while other Unfriendly AIs might be less aggressive than this, it’s hard to imagine an Unfriendly AI that destroys its parent civilization, then sits very quietly doing nothing. It’s even harder to imagine that 999,999,999 out of a billion Unfriendly AIs end up this way.

The Great Filter is not transcendence. Lots of people more enthusiastically propose that the problem isn’t alien species killing themselves, it’s alien species transcending this mortal plane. Once they become sufficiently advanced, they stop being interested in expansion for expansion’s sake. Some of them hang out on their home planet, peacefully cultivating their alien gardens. Others upload themselves to computronium internets, living in virtual reality. Still others become beings of pure energy, doing whatever it is that beings of pure energy do. In any case, they don’t conquer the galaxy or build obvious visible structures.

Which is all nice and well, except what about the Amish aliens? What about the ones who have weird religions telling them that it’s not right to upload their bodies, they have to live in the real world? What about the ones who have crusader religions telling them they have to conquer the galaxy to convert everyone else to their superior way of life? I’m not saying this has to be common. And I know there’s this argument that *advanced* species would be beyond this kind of thing. But man, it only takes one. I can’t believe that not even one in a billion alien civilizations would have some instinctual

preference for galactic conquest for galactic conquest's own sake. I mean, even if most humans upload themselves, there will be a couple who don't and who want to go exploring. You're trying to tell me this model applies to 999,999,999 out of one billion civilizations, and then the very first civilization we test it on, it fails?

The Great Filter is not alien exterminators. It sort of makes sense, from a human point of view. Maybe the first alien species to attain superintelligence was jealous, or just plain jerks, and decided to kill other species before they got the chance to catch up. Knowledgeable people like as Carl Sagan and Stephen Hawking have condemned our reverse-SETI practice of sending messages into space to see who's out there, because everyone out there may be terrible. On this view, the dominant alien civilization is the Great Filter, killing off everyone else while not leaving a visible footprint themselves.

Although I get the precautionary principle, Sagan et al's warnings against sending messages seem kind of silly to me. This isn't a failure to recognize how strong the Great Filter has to be, this is a failure to recognize how powerful a civilization that gets through it can become.

It doesn't matter one way or the other if we broadcast we're here. If there are alien superintelligences out there, *they know*. "Oh, my billion-year-old universe-spanning superintelligence wants to destroy fledgling civilizations, but we just can't find them! If only they would send very powerful radio broadcasts into space so we could figure out where they are!" No. Just no. If there are alien superintelligences out there, they tagged Earth as potential troublemakers sometime in the Cambrian Era and have been watching us very closely ever since. They know what you had for breakfast this morning and they know what Jesus had for breakfast the morning of the Crucifixion.

People worried about accidentally "revealing themselves" to an intergalactic supercivilization are like Sentinel Islanders reluctant to send a message in a bottle lest modern civilization discover their existence – unaware that modern civilization has spy satellites orbiting the planet that can pick out whether or not they shaved that morning.

What about alien exterminators who are okay with weak civilizations, but kill them when they show the first sign of becoming a threat (like inventing fusion power or leaving their home solar system)? Again, you are underestimating billion-year-old universe-spanning superintelligences. Don't flatter yourself here. *You cannot threaten them*.

What about alien exterminators who are okay with weak civilizations, but destroy strong civilizations not because they feel threatened, but just for aesthetic reasons? I can't be certain that's false, but it seems to me that if they have let us continue existing this long, even though we are made of matter that can be used for something else, that has to be a conscious decision made out of something like morality. And because they're omnipotent, they have the ability to satisfy all of their (not logically contradictory) goals at once without worrying about tradeoffs. That makes me think that whatever moral impulse has driven them to allow us to survive will *probably* continue to allow us to survive even if we start annoying them for some reason. When you're omnipotent, the option of stopping the annoyance without harming anyone is just as easy as stopping the annoyance by making everyone involved suddenly vanish.

Three of these four options – x-risk, Unfriendly AI, and alien exterminators – are very very bad for humanity. I think worry about this badness has been a lot of what's driven interest in the Great Filter. I also think these are some of the least likely

possible explanations, which means we should be less afraid of the Great Filter than is generally believed.

Transhumanist Fables

Once upon a time there were three little pigs who went out into the world to build their houses. The first pig was very lazy and built his house out of straw. The second pig was a little harder-working and built his house out of sticks. The third pig was the hardest-working of all, and built his house out of bricks. Then came the Big Bad Wolf. When he saw the house of straw, he huffed and he puffed and he blew the house down, eating the first little pig. When he saw the house of sticks, he huffed and he puffed and he blew the house down, eating the second little pig. When he saw the house of bricks, he got out a bazooka and blew the house to pieces, eating the third little pig.

Moral: Reality doesn't grade on a curve.

Once upon a time there was a big strong troll who lived under a bridge. A little goat went across the bridge, and the troll reached out to grab and eat the goat. "Wait, Mr. Troll!", the goat cried. "Soon my brother is coming, and he is even bigger than I am!" The troll let the goat pass, and soon came another goat, twice as big as the first. The troll reached out to grab and eat him, but the brother likewise objected, saying *his* brother was even bigger. Sure enough, a third goat arrived at the bridge, twice as big as the second, and the troll, now ready for a very hearty dinner, reached out to grab and eat him. "Wait!" said the third goat. "My brother is the biggest of us all!". So the troll let the third goat pass. Then came the fourth goat, who was hundreds of miles tall and blotted out the sun, whose very steps caused earthquakes and made the rivers change course. Without even noticing, he stepped on bridge and troll, pulverizing both to bits.

Moral: Sometimes growth is superexponential.

Once upon a time, Chicken Little ran to her friend Henny Penny. “The sky is falling!” she shouted. “We must tell the king!” Henny Penny joined her, and together they headed toward the capital. On their way they run into their friend Goosey Loosey. “The sky is falling!” they shouted. “We must tell the king!” Goosey Loosey joined them, and together they headed toward the capital. On their way, they ran into the cunning Foxy Loxy. “The sky is falling!” they shouted. “We must tell the king!” “Oh,” said Foxy Loxy. “I know a shortcut to the palace. Follow me into my den.” So the birds all followed Foxy Loxy into his den, where he ate them all, laughing all the while about how gullible they were. Then an asteroid hit Earth, killing everyone.

Moral: Beware the absurdity heuristic.

Once upon a time, a young boy named Jack lived with his mother. Their family was very poor and owned only a single cow. “Go sell this cow at the market,” Jack’s mother told him, “so we will have food to eat for the winter.” Jack went to the market and came back with three beans. “These are magic beans!” he told his mother. “A man told me that when we plant them, they will grow into a beanstalk leading to a land of infinite riches.” His mother pooh – poohed him and threw the beans in the ground angrily. That winter, they both died of hunger.

Moral: Good decision theories should be able to resist Pascal’s Mugging.

Once upon a time, there was an old woodcutter who had no son. He made a little marionette out of pine wood and named it

Pinocchio. Then he wished upon a star that it could become a real boy. The star turned out to be the evil Red Fairy, who brought Pinocchio to life, but told him that if he wanted to be a real boy he must murder everyone in the village. That night, Pinocchio took his father's saw and killed Gepetto and everyone else in town.

Moral: Never create an intelligence unless you are certain it will share your values.

Once upon a time, an evil witch transformed a prince into a frog, telling him that only the kiss of a princess could restore him to his proper form. But although he searched around the world, he could find no princess who was willing to kiss a hideous little frog. Finally, he went to the Wise Wizard. "Gender is a social construct," said the Wise Wizard. "Just declare your gender identity to be female, then kiss yourself on the hand or something." So the frog did that, returned to human form, and ruled the land for many years as a wise and benevolent queen.

Moral: Ability to self-modify is just *ridiculously* powerful.

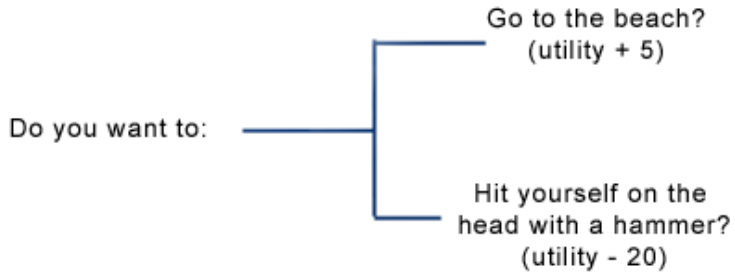
V. Introduction to Game Theory

Backward Reasoning Over Decision Trees

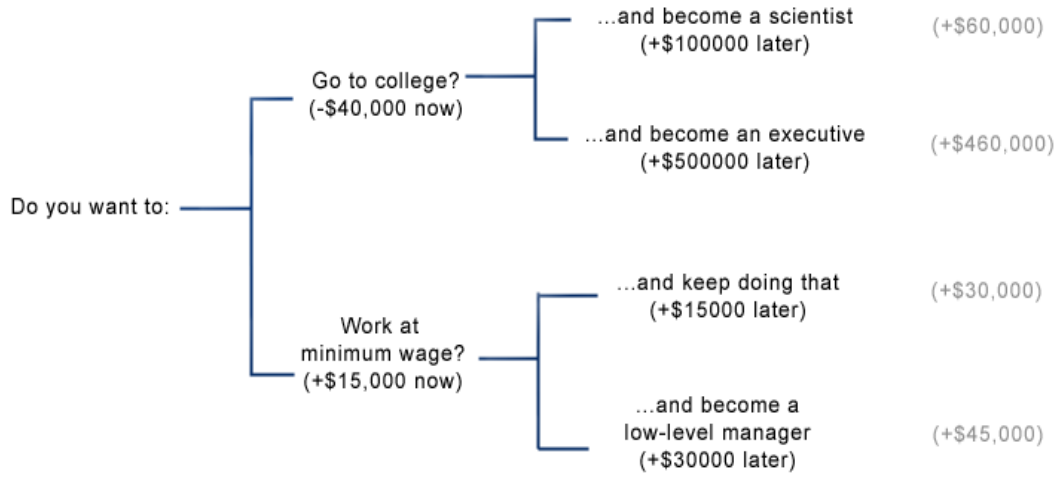
Game theory is the study of how rational actors interact to pursue incentives. It starts with the same questionable premises as economics: that everyone behaves rationally, that everyone is purely self-interested[1], and that desires can be exactly quantified - and uses them to investigate situations of conflict and cooperation.

Here we will begin with some fairly obvious points about decision trees, but by the end we will have the tools necessary to explain a somewhat surprising finding: that giving a US president the additional power of line-item veto may in many cases make the president less able to enact her policies. Starting at the beginning:

The basic unit of game theory is the choice. Rational agents make choices in order to maximize their utility, which is sort of like a measure of how happy they are. In a one-person game, your choices affect yourself and maybe the natural environment, but nobody else. These are pretty simple to deal with:

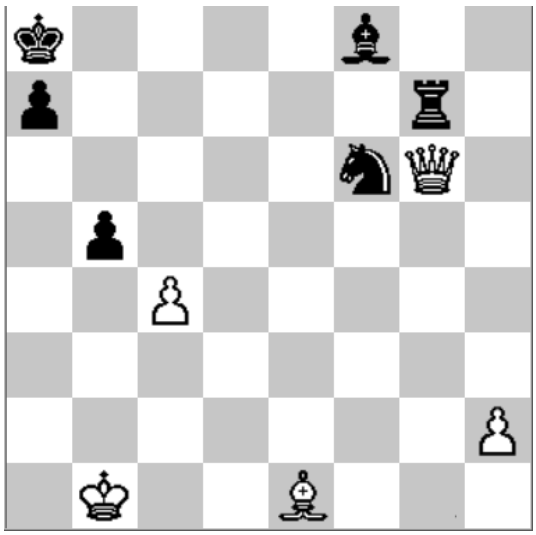


Here we visualize a choice as a branching tree. At each branch, we choose the option with higher utility; in this case, going to the beach. Since each outcome leads to new choices, sometimes the decision trees can be longer than this:

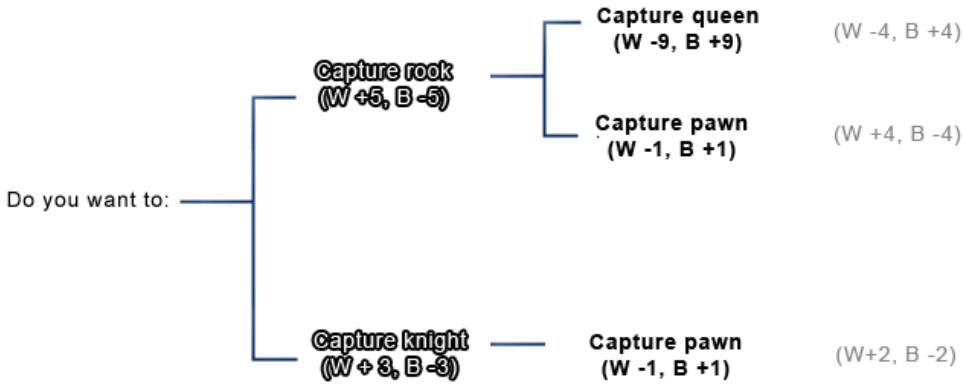


Here's a slightly more difficult decision, denominated in money instead of utility. If you want to make as much money as possible, then your first choice - going to college or starting a minimum wage job right Now - seems to favor the more lucrative minimum wage job. But when you take Later into account, college opens up more lucrative future choices, as measured in the gray totals on the right-hand side. This illustrates the important principle of reasoning backward over decision trees. If you reason forward, taking the best option on the first choice and so on, you end up as a low-level manager. To get the real cash, you've got to start at the end - the total on the right - and then examine what choice at each branch will take you there.

This is all about as obvious as, well, not hitting yourself on the head with a hammer, so let's move on to where it really gets interesting: two-player games.

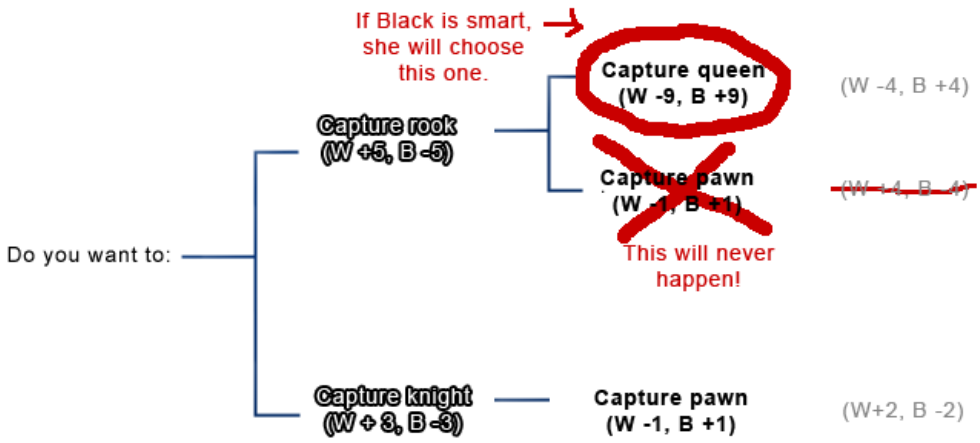


I'm playing White, and it's my move. For simplicity I consider only two options: queen takes knight and queen takes rook. The one chess book I've read values pieces in number of pawns: a knight is worth three pawns, a rook five, a queen nine. So at first glance, it looks like my best move is to take Black's rook. As for Black, I have arbitrarily singled out pawn takes pawn as her preferred move in the current position, but if I play queen takes rook, a new option opens up for her: bishop takes queen. Let's look at the decision tree:



If I foolishly play this two player game the same way I played the one-player go-to-college game, I note that the middle branch has the highest utility for White, so I take the choice that leads there: capture the rook. And then Black plays bishop takes queen, and I am left wailing and gnashing my teeth. What did I do wrong?

I should start by assuming Black will, whenever presented with a choice, take the option with the highest Black utility. Unless Black is stupid, I can cross out any branch that requires Black to play against her own interests. So now the tree looks like this:



The two realistic options are me playing queen takes rook and ending up without a queen and -4 utility, or me playing queen takes knight and ending up with a modest gain of 2 utility.

(my apologies if I've missed some obvious strategic possibility on this particular chessboard; I'm not so good at chess but hopefully the point of the example is clear.)

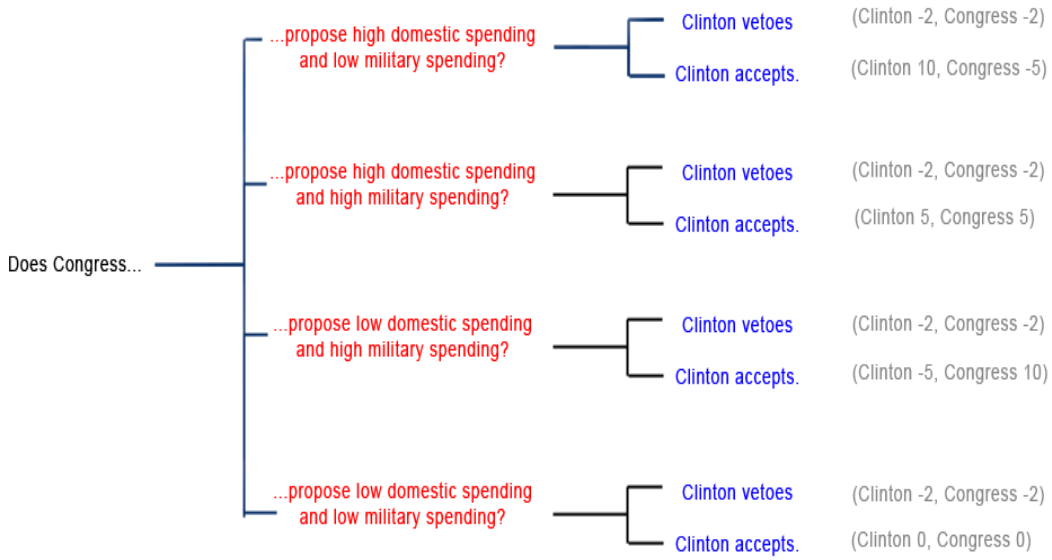
This method of alternating moves in a branching tree matches both our intuitive thought processes during a chess game ("Okay, if I do this, then Black's going to do this, and then I'd do this, and then…") and the foundation of some of the algorithms chess computers like Deep Blue use. In fact, it may seem pretty obvious, or even unnecessary. But it can be used to analyze some more complicated games with counterintuitive results.

Art of Strategy describes a debate from 1990s US politics revolving around so-called "line-item veto" power, the ability to veto only one part of a bill. For example, if Congress passed a bill declaring March to be National Game Theory Month and April to be National Branching Tree Awareness Month, the President could veto only the part about April and leave March intact (as politics currently works, the President could only veto or accept the whole bill). During the '90s, President Clinton fought pretty hard for this power, which seems reasonable as it expands his options when dealing with the hostile Republican Congress.

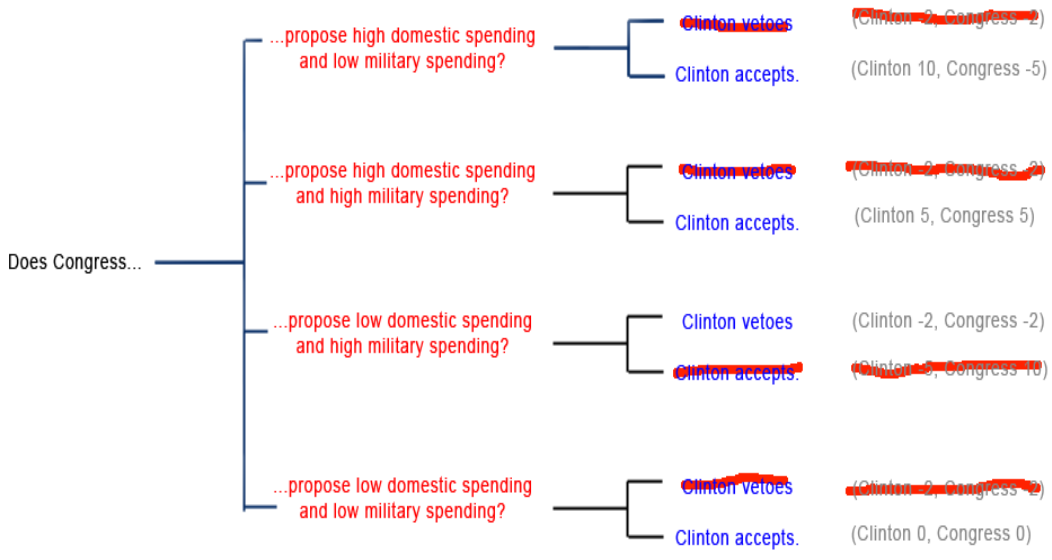
But Dixit and Nalebuff explain that gaining line-item veto powers might hurt a President. How? Branching trees can explain.

Imagine Clinton and the Republican Congress are fighting over a budget. We can think of this as a game of sequential moves, much like chess. On its turn, Congress proposes a budget. On Clinton's turn, he either accepts or rejects the budget. A player "wins" if the budget contains their pet projects. In this game, we start with low domestic and military budgets. Clinton really wants to raise domestic spending (utility +10), and has a minor distaste for raised military spending (utility -5). Congress really wants to raise military spending (utility +10), but has a minor distaste for raised domestic

spending (utility -5). The status quo is zero utility for both parties; if neither party can come to an agreement, voters get angry at them and they both lose 2 utility. Here's the tree when Clinton lacks line-item veto:

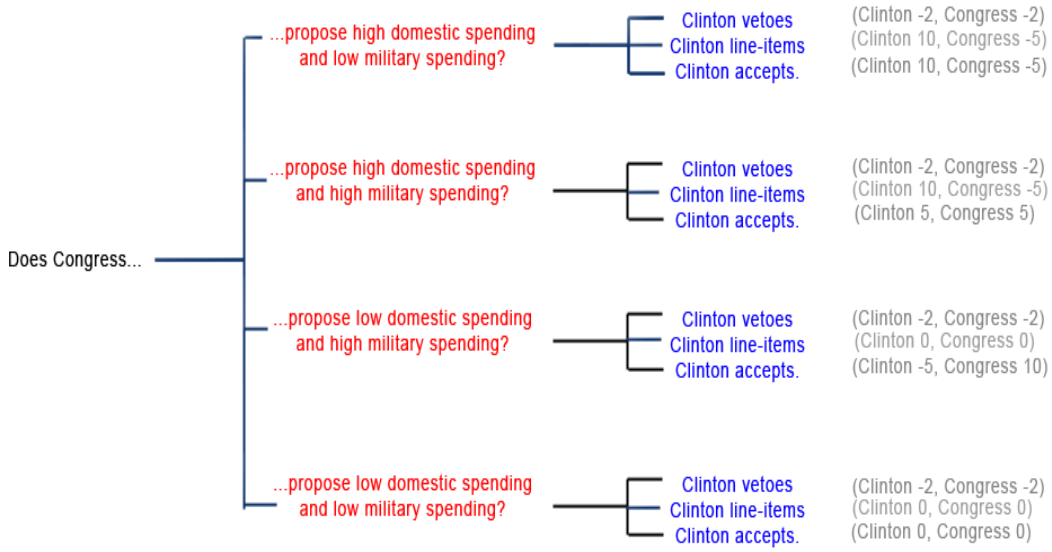


For any particular Republican choice, Clinton will never respond in a way that does not maximize his utility, so the the Republicans reason backward and arrive at something like this:

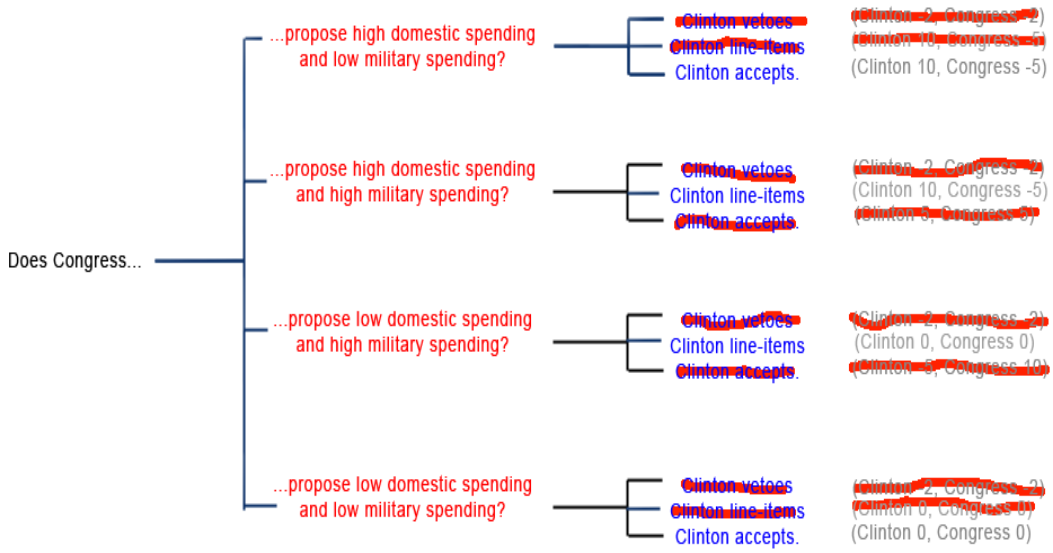


If Republicans are perfectly rational agents, they choose the second option, high domestic and high military spending, to give them their highest plausibly obtainable utility of 5.

But what if Clinton has the line-item veto? Now his options look like this:



If the Republicans stick to their previous choice of "high domestic and high military spending", Clinton line-item vetoes the military spending, and we end up with a situation identical to the first choice: Clinton sitting on a pile of utility, and the Republicans wailing and gnashing their teeth. The Republicans need to come up with a new strategy, and their thought processes, based on Clinton as a utility-maximizer, look like this:



Here Congress's highest utility choice is to propose low domestic spending (it doesn't matter if they give more money to the military or not as this will get line-item vetoed). Let's say they propose low domestic and low military spending, and Clinton accepts. The utilities are (0, 0), and now there is much wailing and gnashing of teeth on both sides (game theorists call this a gnash equilibrium. Maybe you've heard of it.)

But now Clinton has a utility of 0, instead of a utility of 5. Giving him extra options has cost him utility! Why should this happen, and shouldn't he be able to avoid it?

This happened because Clinton's new abilities affect not only his own choices, but those of his opponents (compare Schelling: Strategies of Conflict). He may be able to deal with this if he can make the Republicans trust him.

In summary, simple sequential games can often be explored by reasoning backwards over decision trees representing the choices of the players involved. The next post will discuss simultaneous games and the concept of a Nash equilibrium.

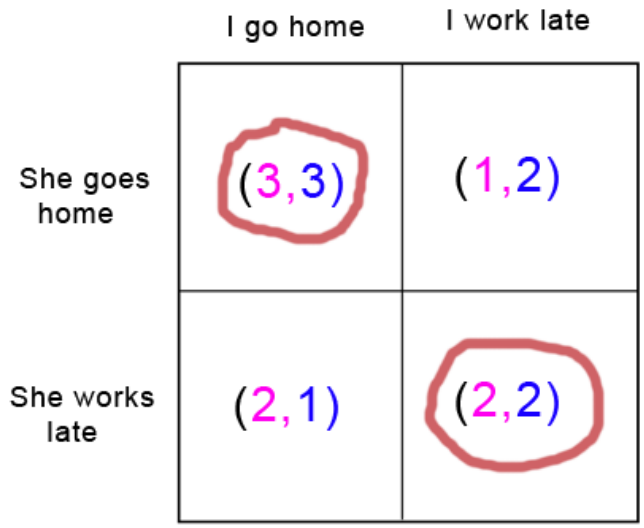
Footnotes:

1: Game theory requires self-interest in that all players' are driven solely by their desire to maximize their own payoff in the game currently being played without regard to the welfare of other players or any external standard of fairness. However, it can also be used to describe the behavior of altruistic agents so long as their altruistic concerns are represented in the evaluation of their payoff.

Nash Equilibria and Schelling Points

A Nash equilibrium is an outcome in which neither player is willing to unilaterally change her strategy, and they are often applied to games in which both players move simultaneously and where decision trees are less useful.

Suppose my girlfriend and I have both lost our cell phones and cannot contact each other. Both of us would really like to spend more time at home with each other (utility 3). But both of us also have a slight preference in favor of working late and earning some overtime (utility 2). If I go home and my girlfriend's there and I can spend time with her, great. If I stay at work and make some money, that would be pretty okay too. But if I go home and my girlfriend's not there and I have to sit around alone all night, that would be the worst possible outcome (utility 1). Meanwhile, my girlfriend has the same set of preferences: she wants to spend time with me, she'd be okay with working late, but she doesn't want to sit at home alone.



This "game" has two Nash equilibria. If we both go home, neither of us regrets it: we can spend time with each other and

we've both got our highest utility. If we both stay at work, again, neither of us regrets it: since my girlfriend is at work, I am glad I stayed at work instead of going home, and since I am at work, my girlfriend is glad she stayed at work instead of going home. Although we both may wish that we had both gone home, neither of us specifically regrets our own choice, given our knowledge of how the other acted.

When all players in a game are reasonable, the (apparently) rational choice will be to go for a Nash equilibrium (why would you want to make a choice you'll regret when you know what the other player chose?) And since John Nash (remember that movie *A Beautiful Mind*?) proved that every game has at least one, all games between well-informed rationalists (who are not also being superrational in a sense to be discussed later) should end in one of these.

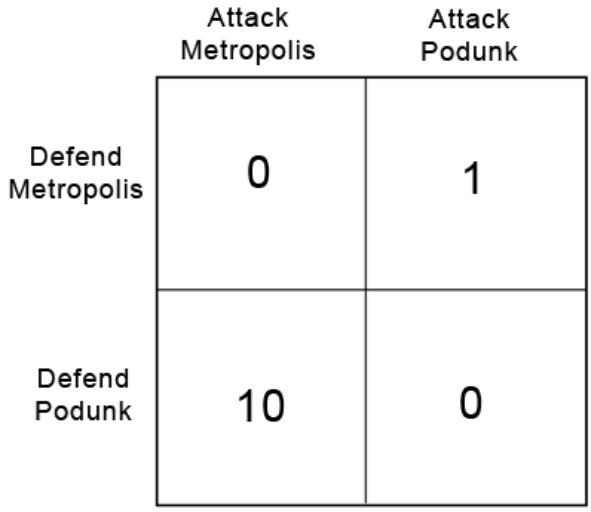
What if the game seems specifically designed to thwart Nash equilibria? Suppose you are a general invading an enemy country's heartland. You can attack one of two targets, East City or West City (you declared war on them because you were offended by their uncreative toponyms). The enemy general only has enough troops to defend one of the two cities. If you attack an undefended city, you can capture it easily, but if you attack the city with the enemy army, they will successfully fight you off.

	Attack East City	Attack West City
Defend East City	(0,1)	(1,0)
Defend West City	(1,0)	(0,1)

Here there is no Nash equilibrium without introducing randomness. If both you and your enemy choose to go to East City, you will regret your choice - you should have gone to West and taken it undefended. If you go to East and he goes to West, he will regret his choice - he should have gone East and stopped you in your tracks. Reverse the names, and the same is true of the branches where you go to West City. So every option has someone regretting their choice, and there is no simple Nash equilibrium. What do you do?

Here the answer should be obvious: it doesn't matter. Flip a coin. If you flip a coin, and your opponent flips a coin, neither of you will regret your choice. Here we see a "mixed Nash equilibrium", an equilibrium reached with the help of randomness.

We can formalize this further. Suppose you are attacking a different country with two new potential targets: Metropolis and Podunk. Metropolis is a rich and strategically important city (utility: 10); Podunk is an out of the way hamlet barely worth the trouble of capturing it (utility: 1).



A so-called first-level player thinks: “Well, Metropolis is a better prize, so I might as well attack that one. That way, if I win I get 10 utility instead of 1”

A second-level player thinks: “Obviously Metropolis is a better prize, so my enemy expects me to attack that one. So if I attack Podunk, he’ll never see it coming and I can take the city undefended.”

A third-level player thinks: “Obviously Metropolis is a better prize, so anyone clever would never do something as obvious as attack there. They’d attack Podunk instead. But my opponent knows that, so, seeking to stay one step ahead of me, he has defended Podunk. He will never expect me to attack Metropolis, because that would be too obvious. Therefore, the city will actually be undefended, so I should take Metropolis.”

And so on ad infinitum, until you become hopelessly confused and have no choice but to spend years developing a resistance to iocane powder.

But surprisingly, there is a single best solution to this problem, even if you are playing against an opponent who, like Professor Quirrell, plays “one level higher than you.”

When the two cities were equally valuable, we solved our problem by flipping a coin. That won't be the best choice this time. Suppose we flipped a coin and attacked Metropolis when we got heads, and Podunk when we got tails. Since my opponent can predict my strategy, he would defend Metropolis every time; I am equally likely to attack Podunk and Metropolis, but taking Metropolis would cost them much more utility. My total expected utility from flipping the coin is 0.5: half the time I successfully take Podunk and gain 1 utility, and half the time I am defeated at Metropolis and gain 0.And this is not a Nash equilibrium: if I had known my opponent's strategy was to defend Metropolis every time, I would have skipped the coin flip and gone straight for Podunk.

So how can I find a Nash equilibrium? In a Nash equilibrium, I don't regret my strategy when I learn my opponent's action. If I can come up with a strategy that pays exactly the same utility whether my opponent defends Podunk or Metropolis, it will have this useful property. We'll start by supposing I am flipping a *biased* coin that lands on Metropolis x percent of the time, and therefore on Podunk (1-x) percent of the time. To be truly indifferent which city my opponent defends, 10x (the utility my strategy earns when my opponent leaves Metropolis undefended) should equal 1(1-x) (the utility my strategy earns when my opponent leaves Podunk undefended). Some quick algebra finds that 10x = 1(1-x) is satisfied by x = 1/11. So I should attack Metropolis 1/11 of the time and Podunk 10/11 of the time.

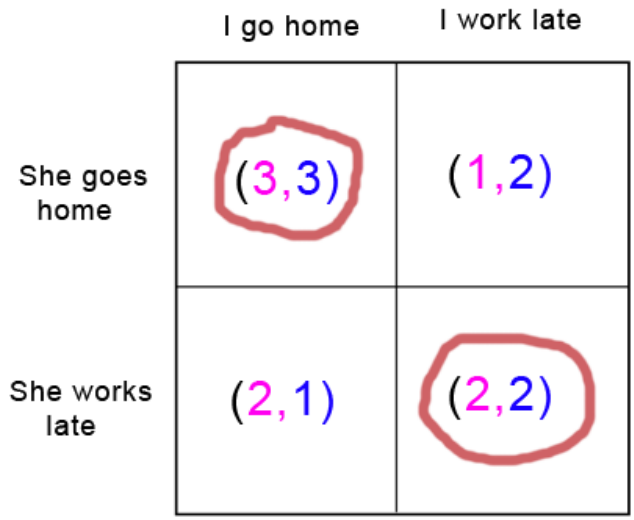
My opponent, going through a similar process, comes up with the suspiciously similar result that he should defend Metropolis 10/11 of the time, and Podunk 1/11 of the time.

If we both pursue our chosen strategies, I gain an average 0.9090… utility each round, soundly beating my previous

record of 0.5, and my opponent suspiciously loses an average -.9090 utility. It turns out there is no other strategy I can use to consistently do better than this when my opponent is playing optimally, and that even if I knew my opponent's strategy I would not be able to come up with a better strategy to beat it. It also turns out that there is no other strategy my opponent can use to consistently do better than this if I am playing optimally, and that my opponent, upon learning my strategy, doesn't regret his strategy either.

In *The Art of Strategy*, Dixit and Nalebuff cite a real-life application of the same principle in, of all things, penalty kicks in soccer. A right-footed kicker has a better chance of success if he kicks to the right, but a smart goalie can predict that and will defend to the right; a player expecting this can accept a less spectacular kick to the left if he thinks the left will be undefended, but a very smart goalie can predict this too, and so on. Economist Ignacio Palacios-Huerta laboriously analyzed the success rates of various kickers and goalies on the field, and found that they actually pursued a mixed strategy generally within 2% of the game theoretic ideal, proving that people are pretty good at doing these kinds of calculations unconsciously.

So every game really does have at least one Nash equilibrium, even if it's only a mixed strategy. But some games can have many, many more. Recall the situation between me and my girlfriend:



There are two Nash equilibria: both of us working late, and both of us going home. If there were only one equilibrium, and we were both confident in each other's rationality, we could choose that one and there would be no further problem. But in fact this game does present a problem: intuitively it seems like we might still make a mistake and end up in different places.

Here we might be tempted to just leave it to chance; after all, there's a 50% probability we'll both end up choosing the same activity. But other games might have thousands or millions of possible equilibria and so will require a more refined approach.

Art of Strategy describes a game show in which two strangers were separately taken to random places in New York and promised a prize if they could successfully meet up; they had no communication with one another and no clues about how such a meeting was to take place. Here there are a nearly infinite number of possible choices: they could both meet at the corner of First Street and First Avenue at 1 PM, they could both meet at First Street and Second Avenue at 1:05 PM, etc. Since neither party would regret their actions (if I went to First and First at 1 and found you there, I would be thrilled) these are all Nash equilibria.

Despite this mind-boggling array of possibilities, in fact all six episodes of this particular game ended with the two contestants meeting successfully after only a few days. The most popular meeting site was the Empire State Building at noon.

How did they do it? The world-famous Empire State Building is what game theorists call focal: it stands out as a natural and obvious target for coordination. Likewise noon, classically considered the very middle of the day, is a focal point in time. These focal points, also called Schelling points after theorist Thomas Schelling who discovered them, provide an obvious target for coordination attempts.

What makes a Schelling point? The most important factor is that it be special. The Empire State Building, depending on when the show took place, may have been the tallest building in New York; noon is the only time that fits the criteria of "exactly in the middle of the day", except maybe midnight when people would be expected to be too sleepy to meet up properly.

Of course, specialness, like beauty, is in the eye of the beholder. David Friedman writes:

> *Two people are separately confronted with the list of numbers [2, 5, 9, 25, 69, 73, 82, 96, 100, 126, 150] and offered a reward if they independently choose the same number. If the two are mathematicians, it is likely that they will both choose 2—the only even prime. Non-mathematicians are likely to choose 100—a number which seems, to the mathematicians, no more unique than the other two exact squares. Illiterates might agree on 69, because of its peculiar symmetry—as would, for a*

> *different reason, those whose interest in numbers is more prurient than mathematical.*

A recent open thread comment pointed out that you can justify anything with "for decision-theoretic reasons" or "due to meta-level concerns". I humbly propose adding "as a Schelling point" to this list, except that the list is tongue-in-cheek and Schelling points really do explain almost everything - stock markets, national borders, marriages, private property, religions, fashion, political parties, peace treaties, social networks, software platforms and languages all involve or are based upon Schelling points. In fact, whenever something has "symbolic value" a Schelling point is likely to be involved in some way. I hope to expand on this point a bit more later.

Sequential games can include one more method of choosing between Nash equilibria: the idea of a subgame-perfect equilibrium, a special kind of Nash equlibrium that remains a Nash equilibrium for every subgame of the original game. In more intuitive terms, this equilibrium means that even in a long multiple-move game no one at any point makes a decision that goes against their best interests (remember the example from the last post, where we crossed out the branches in which Clinton made implausible choices that failed to maximize his utility?) Some games have multiple Nash equilibria but only one subgame-perfect one; we'll examine this idea further when we get to the iterated prisoners' dilemma and ultimatum game.

In conclusion, every game has at least one Nash equilibrium, a point at which neither player regrets her strategy even when she knows the other player's strategy. Some equilibria are simple choices, others involve plans to make choices randomly according to certain criteria. Purely rational players will

always end up at a Nash equilibrium, but many games will have multiple possible equilibria. If players are trying to coordinate, they may land at a Schelling point, an equilibria which stands out as special in some way.

Introduction to Prisoners' Dilemma

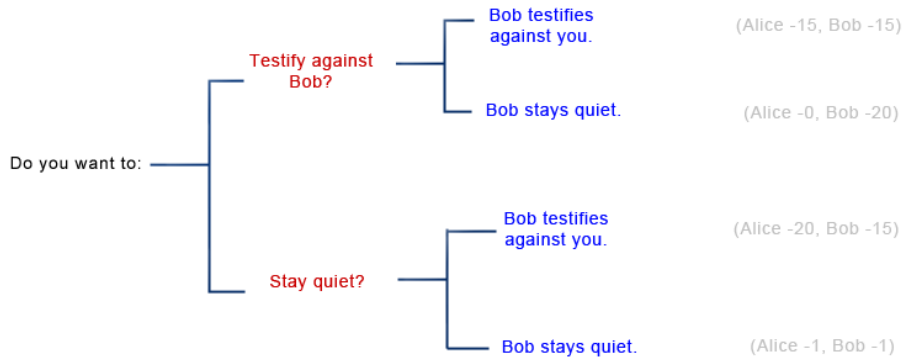
Related to: Previous posts on the Prisoners' Dilemma

Sometimes Nash equilibria just don't match our intuitive criteria for a good outcome. The classic example is the Prisoners' Dilemma.

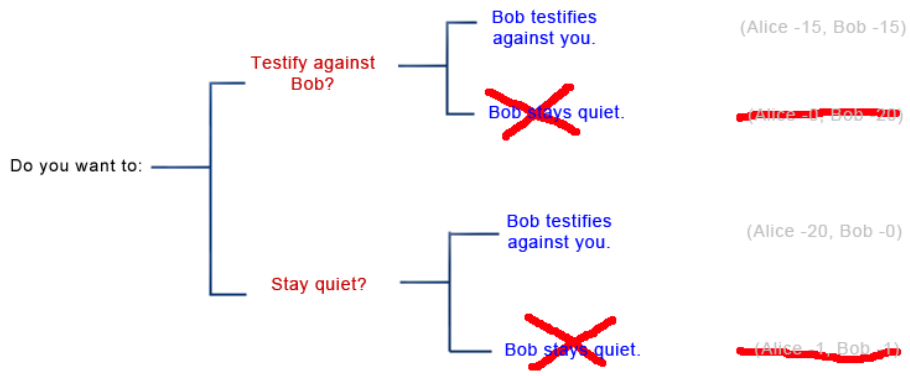
The police arrest two criminals, Alice and Bob, on suspicion of murder. The police admit they don't have enough evidence to convict the pair of murder, but they do have enough evidence to convict them of a lesser offence, possession of a firearm. They place Alice and Bob in separate cells and offer them the following deal:

"If neither of you confess, we'll have to charge you with possession, which will land you one year in jail. But if you turn state's witness against your partner, we can convict your partner of murder and give her the full twenty year sentence; in exchange, we will let you go free. Unless, that is, both of you testify against each other; in that case, we'll give you both fifteen years."

Alice's decision tree looks like this (note that although Alice and Bob make their decisions simultaneously, I've represented it with Alice's decision first, which is a little sketchy but should illustrate the point):



If we use the same strategy we used as a chess player, we can cross out options where Bob decides to spend extra years in jail for no personal benefit, and we're left with this:



Seen like this, the choice is clear. If you stay quiet (“cooperate”), Bob turns on you, and you are left in jail alone for twenty years, wailing and gnashing your teeth. So instead, you both turn on each other (“defect”), and end up with a sentence barely any shorter.

Another way to “prove” that defection is the “right” choice places Bob’s decision first. What if you knew Bob would choose to cooperate with you? Then your choice would be between defecting and walking free, or cooperating and spending a whole year in jail - here defection wins out. But what if you knew Bob would choose to defect against you? Then your choice would be between defecting and losing fifteen years, or cooperating and losing twenty - again defection wins out. Since Bob can only either defect or cooperate, and since defection is better in both branches, “clearly” defection is the best option.

But a lot of things about this solution seem intuitively stupid. For example, when Bob goes through the same reasoning, your “rational” solution ends up with both of you in jail for fifteen years, but if you had both cooperated, you would have been out after a year. Both cooperating is better for both of you than both defecting, but you still both defect.

And if you still don’t find that odd, imagine a different jurisdiction where the sentence for possession is only one day, and the police will only take a single day off your sentence for testifying against an accomplice. Now a pair of cooperators would end up with only a day in jail each, and a pair of defectors would end up with nineteen years, three hundred sixty four days each. Yet the math still tells you to defect!

Unfortunately, your cooperation only helps Bob, and Bob's cooperation only helps you. We can think of the Prisoner's Dilemma as a problem: both you and Bob prefer (cooperate, cooperate) to (defect, defect), but as it is, you're both going to end out with (defect, defect) and it doesn't seem like there's much you can do about it. To "solve" the Prisoner's Dilemma would be to come up with a way to make you and Bob pick the more desirable (cooperate, cooperate) outcome.

One proposed solution to the Prisoner's Dilemma is to iterate it - to assume it will happen multiple times in succession, as if Alice and Bob are going to commit new crimes as soon as they both get out of prison. In this case, you can threaten to reciprocate; to promise to reward cooperation with future cooperation and punish defection with future defection. Suppose Alice and Bob plan to commit two crimes, and before the first crime both promise to stay quiet on the second crime if and only if their partner stays quiet on the first. Now your decision tree as Alice looks like this:

And your calculation of Bob's thought processes go like this:

Remember that, despite how the graph looks, your first choice and Bob's first choice are simultaneous: they can't causally affect each other. So as Alice, you reason like this: On the top, Bob knows that if you testify against him, his choice will be either to testify against you (leading to the branch where you both testify against each other again next time) or to stay quiet (leading to a branch where next time he testifies against you but you stay quiet). So Bob reasons that if you testify against him, he should stay quiet this time.

On the bottom, Bob knows that if you don't testify against him, he can either testify against you (leading to the branch where you testify against him next time but he stays quiet) or stay quiet (leading to the branch where you both stay quiet again next time). Therefore, if you don't testify against him, Bob won't testify against you.

So you know that no matter what you do this time, Bob won't testify against you. That means your choice is between branches 2 and 4: Bob testifying against you next time or Bob not testifying against you next time. You prefer Branch 4, so you decide not to testify against Bob. The dilemma ends with neither of you testifying against each other in either crime, and both of you getting away with very light two year sentences.

The teeny tiny little flaw in this plan is that Bob may be a dirty rotten liar. Maybe he says he'll reciprocate, and so you both stay quiet after the first crime. Upon getting out of jail you continue your crime spree, predictably get re-arrested, and you stay quiet like you

said you would to reward Bob's cooperation last time. But at the trial, you get a nasty surprise: Bob defects against you and walks free, and you end up with a twenty year sentence.

If we ratchet up to sprees of one hundred crimes and subsequent sentences (presumably committed by immortal criminals who stubbornly refuse to be cowed by the police's 100% conviction rate) on first glance it looks like we can successfully ensure Bob's cooperation on 99 of those crimes. After all, Bob won't want to defect on crime 50, because I could punish him on crime 51. He won't want to defect on crime 99, because I could punish him on crime 100. But he *will* want to defect on crime 100, because he gains either way and there's nothing I can do to punish him.

Here's where it gets weird. I assume Bob is a rational utility-maximizer and so will always defect on crime 100, since it benefits him and I can't punish him for it. So since I'm also rational, I might as well also defect on crime 100; my previous incentive to cooperate was to ensure Bob's good behavior, but since Bob won't show good behavior on crime 100 no matter what I do, I might as well look after my own interests.

But if we both know that we're both going to defect on crime 100 no matter what, then there's no incentive to cooperate on crime 99. After all, the only incentive to cooperate on crime 99 was to ensure my rival's cooperation on crime 100, and since that's out of the picture anyway, I might as well shorten my sentence a little.

Sadly, this generalizes by a sort of proof by induction. If crime N will always be (defect, defect), then crime N-1 should also always be (defect, defect), which means we should defect on all of the hundred crimes in our spree.

This feat of reasoning has limited value in real life, where perfectly rational immortal criminals rarely plot in smoke-filled rooms to commit exactly one hundred crimes together; criminals who are uncertain exactly when their crime sprees will come to a close still have incentive to cooperate. But it still looks like we're going to

need a better solution than simply iterating the dilemma. The next post will discuss possibilities for such a solution.

Real-World Solutions to Prisoners' Dilemmas

Why should there be real world solutions to Prisoners' Dilemmas? Because such dilemmas are a real-world problem.

If I am assigned to work on a school project with a group, I can either cooperate (work hard on the project) or defect (slack off while reaping the rewards of everyone else's hard work). If everyone defects, the project doesn't get done and we all fail - a bad outcome for everyone. If I defect but you cooperate, then I get to spend all day on the beach and still get a good grade - the best outcome for me, the worst for you. And if we all cooperate, then it's long hours in the library but at least we pass the class - a "good enough" outcome, though not quite as good as me defecting against everyone else's cooperation. This exactly mirrors the Prisoner's Dilemma.

Diplomacy - both the concept and the board game - involves Prisoners' Dilemmas. Suppose Ribbentrop of Germany and Molotov of Russia agree to a peace treaty that demilitarizes their mutual border. If both cooperate, they can move their forces to other theaters, and have moderate success there - a good enough outcome. If Russia cooperates but Germany defects, it can launch a surprise attack on an undefended Russian border and enjoy spectacular success there (for a while, at least!) - the best outcome for Germany and the worst for Russia. But if both defect, then neither has any advantage at the German-Russian border, and they lose the use of those troops in other theaters as well - a bad outcome for both. Again, the Prisoner's Dilemma.

Civilization - again, both the concept and the game - involves Prisoners' Dilemmas. If everyone follows the rules and creates a stable society (cooperates), we all do pretty well. If everyone else works hard and I turn barbarian and pillage you (defect), then I get all of your stuff without having to work for it and you get nothing - the best solution for me, the worst for you. If everyone becomes a

barbarian, there's nothing to steal and we all lose out. Prisoner's Dilemma.

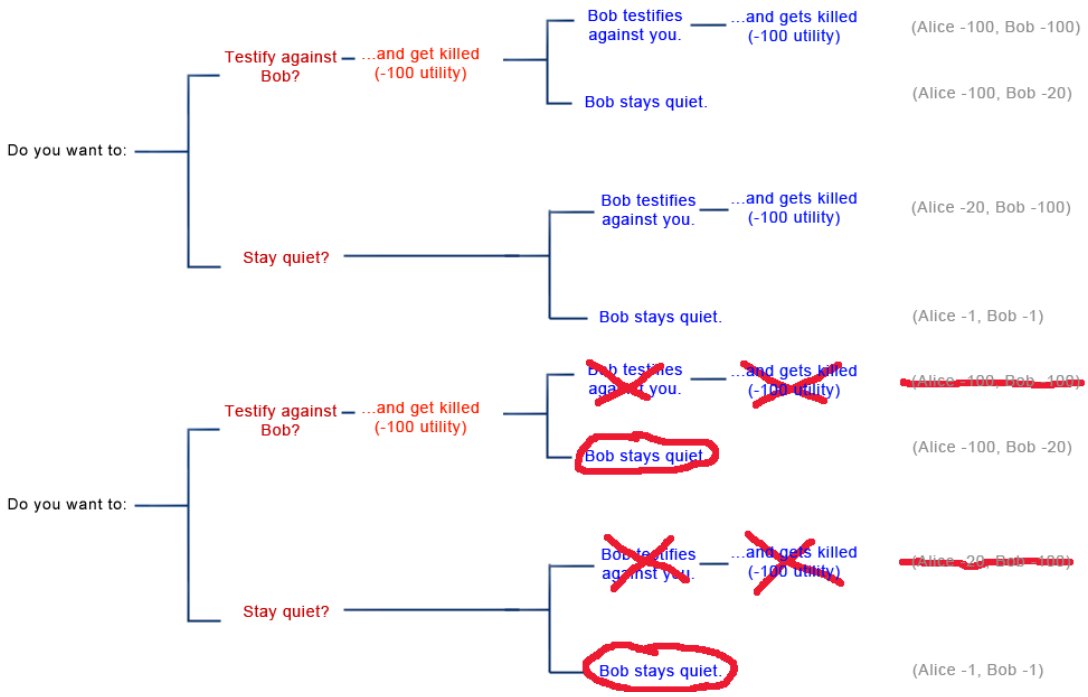
If everyone who worries about global warming cooperates in cutting emissions, climate change is averted and everyone is moderately happy. If everyone else cooperates in cutting emissions, but one country defects, climate change is still mostly averted, and the defector is at a significant economic advantage. If everyone defects and keeps polluting, the climate changes and everyone loses out. Again a Prisoner's Dilemma,

Prisoners' Dilemmas even come up in nature. In baboon tribes, when a female is in "heat", males often compete for the chance to woo her. The most successful males are those who can get a friend to help fight off the other monkeys, and who then helps that friend find his own monkey loving. But these monkeys are tempted to take their friend's female as well. Two males who cooperate each seduce one female. If one cooperates and the other defects, he has a good chance at both females. But if the two can't cooperate at all, then they will be beaten off by other monkey alliances and won't get to have sex with anyone. Still a Prisoner's Dilemma!

So one might expect the real world to have produced some practical solutions to Prisoners' Dilemmas.

One of the best known such systems is called "society". You may have heard of it. It boasts a series of norms, laws, and authority figures who will punish you when those norms and laws are broken.

Imagine that the two criminals in the original example were part of a criminal society - let's say the Mafia. The Godfather makes Alice and Bob an offer they can't refuse: turn against one another, and they will end up "sleeping with the fishes" (this concludes my knowledge of the Mafia). Now the incentives are changed: defecting against a cooperator doesn't mean walking free, it means getting murdered.



Both prisoners cooperate, and amazingly the threat of murder ends up making them both better off (this is also the gist of some of the strongest arguments against libertarianism: in Prisoner's Dilemmas, threatening force against rational agents can increase the utility of all of them!)

Even when there is no godfather, society binds people by concern about their "reputation". If Bob got a reputation as a snitch, he might never be able to work as a criminal again. If a student gets a reputation for slacking off on projects, she might get ostracized on the playground. If a country gets a reputation for backstabbing, others might refuse to make treaties with them. If a person gets a reputation as a bandit, she might incur the hostility of those around her. If a country gets a reputation for not doing enough to fight global warming, it might…well, no one ever said it was a perfect system.

Aside from humans in society, evolution is also strongly motivated to develop a solution to the Prisoner's Dilemma. The Dilemma troubles not only lovestruck baboons, but ants, minnows, bats, and even viruses. Here the payoff is denominated not in years of jail time, nor in dollars, but in reproductive fitness and number of potential offspring - so evolution will certainly take note.

Most people, when they hear the rational arguments in favor of defecting every single time on the iterated 100-crime Prisoner's Dilemma, will feel some kind of emotional resistance. Thoughts like "Well, maybe I'll try cooperating anyway a few times, see if it works", or "If I promised to cooperate with my opponent, then it would be dishonorable for me to defect on the last turn, even if it helps me out., or even "Bob is my friend! Think of all the good times we've had together, robbing banks and running straight into waiting police cordons. I could never betray him!"

And if two people with these sorts of emotional hangups play the Prisoner's Dilemma together, they'll end up cooperating on all hundred crimes, getting out of jail in a mere century and leaving rational utility maximizers to sit back and wonder how they did it.

Here's how: imagine you are a supervillain designing a robotic criminal (who's that go-to supervillain Kaj always uses for situations like this? Dr. Zany? Okay, let's say you're him). You expect to build several copies of this robot to work as a team, and expect they might end up playing the Prisoner's Dilemma against each other. You want them out of jail as fast as possible so they can get back to furthering your nefarious plots. So rather than have them bumble through the whole rational utility maximizing thing, you just insert an extra line of code: "in a Prisoner's Dilemma, always cooperate with other robots". Problem solved.

Evolution followed the same strategy (no it didn't; this is a massive oversimplification). The emotions we feel around friendship, trust, altruism, and betrayal are partly a built-in hack to succeed in cooperating on Prisoner's Dilemmas where a rational utility-maximizer would defect a hundred times and fail miserably. The evolutionarily dominant strategy is commonly called "Tit-for-tat" - basically, cooperate if and only if your opponent did so last time.

This so-called "superrationality" appears even more clearly in the Ultimatum Game. Two players are given $100 to distribute among themselves in the following way: the first player proposes a distribution (for example, "Fifty for me, fifty for you") and then the

second player either accepts or rejects the distribution. If the second player accepts, the players get the money in that particular ratio. If the second player refuses, no one gets any money at all.

The first player's reasoning goes like this: "If I propose $99 for myself and $1 for my opponent, that means I get a lot of money and my opponent still has to accept. After all, she prefers $1 to $0, which is what she'll get if she refuses.

In the Prisoner's Dilemma, when players were able to communicate beforehand they could settle upon a winning strategy of precommiting to reciprocate: to take an action beneficial to their opponent if and only if their opponent took an action beneficial to them. Here, the second player should consider the same strategy: precommit to an ultimatum (hence the name) that unless Player 1 distributes the money 50-50, she will reject the offer.

But as in the Prisoner's Dilemma, this fails when you have no reason to expect your opponent to follow through on her precommitment. Imagine you're Player 2, playing a single Ultimatum Game against an opponent you never expect to meet again. You dutifully promise Player 1 that you will reject any offer less than 50-50. Player 1 offers 80-20 anyway. You reason "Well, my ultimatum failed. If I stick to it anyway, I walk away with nothing. I might as well admit it was a good try, give in, and take the $20. After all, rejecting the offer won't magically bring my chance at $50 back, and there aren't any other dealings with this Player 1 guy for it to influence."

This is seemingly a rational way to think, but if Player 1 knows you're going to think that way, she offers 99-1, same as before, no matter how sincere your ultimatum sounds.

Notice all the similarities to the Prisoner's Dilemma: playing as a "rational economic agent" gets you a bad result, it looks like you can escape that bad result by making precommitments, but since the other player can't trust your precommitments, you're right back where you started

If evolutionary solutions to the Prisoners' Dilemma look like trust or friendship or altruism, solutions to the Ultimatum Game involve different emotions entirely. The Sultan presumably does not want you to elope with his daughter. He makes an ultimatum: "Touch my daughter, and I will kill you." You elope with her anyway, and when his guards drag you back to his palace, you argue: "Killing me isn't going to reverse what happened. Your ultimatum has failed. All you can do now by beheading me is get blood all over your beautiful palace carpet, which hurts you as well as me - the equivalent of pointlessly passing up the last dollar in an Ultimatum Game where you've just been offered a 99-1 split."

The Sultan might counter with an argument from social institutions: "If I let you go, I will look dishonorable. I will gain a reputation as someone people can mess with without any consequences. My choice isn't between bloody carpet and clean carpet, it's between bloody carpet and people respecting my orders, or clean carpet and people continuing to defy me."

But he's much more likely to just shout an incoherent stream of dreadful Arabic curse words. Because just as friendship is the evolutionary solution to a Prisoner's Dilemma, so anger is the evolutionary solution to an Ultimatum Game. As various gurus and psychologists have observed, anger makes us irrational. But this is the good kind of irrationality; it's the kind of irrationality that makes us pass up a 99-1 split even though the decision costs us a dollar.

And if we know that humans are the kind of life-form that tends to experience anger, then if we're playing an Ultimatum Game against a human, and that human precommits to rejecting any offer less than 50-50, we're much more likely to believe her than if we were playing against a rational utility-maximizing agent - and so much more likely to give the human a fair offer.

It is distasteful and a little bit contradictory to the spirit of rationality to believe it should lose out so badly to simple emotion, and the problem might be correctable. Here we risk crossing the poorly charted border between game theory and decision theory and

reaching ideas like timeless decision theory: that one should act as if one's choices determined the output of the algorithm one instantiates (or more simply, you should assume everyone like you will make the same choice you do, and take that into account when choosing.)

More practically, however, most real-world solutions to Prisoner's Dilemmas and Ultimatum Games still hinge on one of three things: threats of reciprocation when the length of the game is unknown, social institutions and reputation systems that make defection less attractive, and emotions ranging from cooperation to anger that are hard-wired into us by evolution. In the next post, we'll look at how these play out in practice.

Interlude for Behavioral Economics

The so-called "rational" solutions to the Prisoners' Dilemma and Ultimatum Game are suboptimal to say the least. Humans have various kludges added by both nature or nurture to do better, but they're not perfect and they're certainly not simple. They leave entirely open the question of what real people will actually do in these situations, a question which can only be addressed by hard data.

As in so many other areas, our most important information comes from reality television. *The Art of Strategy* discusses a US game show "Friend or Foe" where a team of two contestants earned money by answering trivia questions. At the end of the show, the team used a sort-of Prisoner's Dilemma to split their winnings: each team member chose "Friend" (cooperate) or "Foe" (defect). If one player cooperated and the other defected, the defector kept 100% of the pot. If both cooperated, each kept 50%. And if both defected, neither kept anything (this is a significant difference from the standard dilemma, where a player is a little better off defecting than cooperating if her opponent defects).

Players chose "Friend" about 45% of the time. Significantly, this number remained constant despite the size of the pot: they were no more likely to cooperate when splitting small amounts of money than large.

Players seemed to want to play "Friend" if and only if they expected their opponents to do so. This is not rational, but it accords with the "Tit-for-Tat" strategy hypothesized to be the evolutionary solution to Prisoner's Dilemma. This played out on the show in a surprising way: players' choices started off random, but as the show went on and contestants began

participating who had seen previous episodes, they began to base their decision on observable characteristics about their opponents. For example, in the first season women cooperated more often than men, so by the second season a player was cooperating more often if their opponent was a woman - whether or not that player was a man or woman themselves.

Among the superficial characteristics used, the only one to reach statistical significance according to the study was age: players below the median age of 27 played “Foe” more often than those over it (65% vs. 39%, $p < .001$). Other nonsignificant tendencies were for men to defect more than women (53% vs. 46%, $p=.34$) and for black people to defect more than white people (58% vs. 48%, $p=.33$). These nonsignificant tendencies became important because the players themselves attributed significance to them: for example, by the second season women were playing “Foe” 60% of the time against men but only 45% of the time against women ($p<.01$) presumably because women were perceived to be more likely to play “Friend” back; also during the second season, white people would play “Foe” 75% against black people, but only 54% of the time against other white people.

(This risks self-fulfilling prophecies. If I am a black man playing a white woman, I expect she will expect me to play “Foe” against her, and she will “reciprocate” by playing “Foe” herself. Therefore, I may choose to “reciprocate” against her by playing “Foe” myself, even if I wasn’t originally intending to do so, and other white women might observe this, thus creating a vicious cycle.)

In any case, these attempts at coordinated play worked, but only imperfectly. By the second season, 57% of pairs chose the same option - either (C, C) or (D, D).

Art of Strategy included another great Prisoner's Dilemma experiment. In this one, the experimenters spoiled the game: they told both players that they would be deciding simultaneously, but in fact, they let Player 1 decide first, and then secretly approached Player 2 and told her Player 1's decision, letting Player 2 consider this information when making her own choice.

Why should this be interesting? From the previous data, we know that humans play "tit-for-expected-tat": they will generally cooperate if they believe their opponent will cooperate too. We can come up with two hypotheses to explain this behavior. First, this could be a folk version of Timeless Decision Theory or Hofstadter's superrationality; a belief that their own decision literally determines their opponent's decision. Second, it could be based on a belief in fairness: if I think my opponent cooperated, it's only decent that I do the same.

The "researchers spoil the setup" experiment can distinguish between these two hypotheses. If people believe their choice determines that of their opponent, then once they know their opponent's choice they no longer have to worry and can freely defect to maximize their own winnings. But if people want to cooperate to reward their opponent, then learning that their opponent cooperated for sure should only increase their willingness to reciprocate.

The results: If you tell the second player that the first player defected, 3% still cooperate (apparently 3% of people are Jesus). If you tell the second player that the first player cooperated……only 16% cooperate. When the same researchers in the same lab didn't tell the second player anything, 37% cooperated.

This is a pretty resounding victory for the "folk version of superrationality" hypothesis. 21% of people wouldn't cooperate if they heard their opponent defected, wouldn't cooperate if they heard their opponent cooperated, but will cooperate if they don't know which of those two their opponent played.

Moving on to the Ultimatum Game: very broadly, the first player usually offers between 30 and 50 percent, and the second player tends to accept. If the first player offers less than about 20 percent, the second player tends to reject it.

Like the Prisoner's Dilemma, the amount of money at stake doesn't seem to matter. This is really surprising! Imagine you played an Ultimatum Game for a billion dollars. The first player proposes $990 million for herself, $10 million for you. On the one hand, this is a 99-1 split, just as unfair as $99 versus $1. On the other hand, ten million dollars!

Although tycoons have yet to donate a billion dollars to use for Ultimatum Game experiments, researchers have done the next best thing and flown out to Third World countries where even $100 can be an impressive amount of money. In games in Indonesia played for a pot containing a sixth of Indonesians' average yearly income, Indonesians still rejected unfair offers. In fact, at these levels the first player tended to propose fairer deals than at lower stakes - maybe because it would be a disaster if her offer get rejected.

It was originally believed that results in the Ultimatum Game were mostly independent of culture. Groups in the US, Israel, Japan, Eastern Europe, and Indonesia all got more or less the same results. But this elegant simplicity was, like so many other things, ruined by the Machiguenga Indians of eastern

Peru. They tend to make offers around 25%, and will accept pretty much anything.

One more interesting finding: people who accept low offers in the Ultimatum Game have lower testosterone than those who reject them.

There is a certain degenerate form of the Ultimatum Game called the Dictator Game. In the Dictator Game, the second player doesn't have the option of vetoing the first player's distribution. In fact, the second player doesn't do anything at all; the first player distributes the money, both players receive the amount of money the first player decided upon, and the game ends. A perfectly selfish first player would take 100% of the money in the Dictator Game, leaving the second player with nothing.

In a metaanalysis of 129 papers consisting of over 41,000 individual games, the average amount the first player gave the second player was 28.35%. 36% of first players take everything, 17% divide the pot equally, and 5% give everything to the second player, nearly doubling our previous estimate of what percent of people are Jesus.

The meta-analysis checks many different results, most of which are insignificant, but a few stand out. Subjects playing the dictator game "against" a charity are much more generous; up to a quarter give everything. When the experimenter promises to "match" each dollar given away (eg the dictator gets $100, but if she gives it to the second player the second player gets $200), the dictator gives much more (somewhat surprising, as this might be an excuse to keep $66 for yourself and get away with it by claiming that both players still got equal money). On the other hand, if the experimenters give the second player a free $100, so that they start off richer than the

dictator, the dictator compensates by not giving them nearly as much money.

Old people give more than young people, and non-students give more than students. People from “primitive” societies give more than people from more developed societies, and the more primitive the society, the stronger the effect. The most important factor, though? As always, sex. Women both give more and get more in dictator games.

It is somewhat inspiring that so many people give so much in this game, but before we become too excited about the fundamental goodness of humanity, Art of Strategy mentions a great experiment by Dana, Cain, and Dawes. The subjects were offered a choice: either play the Dictator Game with a second player for $10, or get $9 and the second subject is sent home and never even knows what the experiment is about. A third of participants took the second option.

So generosity in the Dictator Game isn’t always about wanting to help other people. It seems to be about knowing, deep down, that some anonymous person who probably doesn’t even know your name and who will never see you again is disappointed in you. Remove the little problem of the other person knowing what you did, and they will not only keep the money, but even be willing to pay the experiment a dollar to keep them quiet.

What is Signaling, Really?

The most commonly used introduction to signaling, promoted both by Robin Hanson and in *The Art of Strategy*, starts with college degrees. Suppose, there are two kinds of people, smart people and stupid people; and suppose, with wild starry-eyed optimism, that the populace is split 50-50 between them. Smart people would add enough value to a company to be worth a $100,000 salary each year, but stupid people would only be worth $40,000. And employers, no matter how hard they try to come up with silly lateral-thinking interview questions like "How many ping-pong balls could fit in the Sistine Chapel?", can't tell the difference between them.

Now suppose a certain college course, which costs $50,000, passes all smart people but flunks half the stupid people. A strategic employer might declare a policy of hiring (for a one year job; let's keep this model simple) graduates at $100,000 and non-graduates at $40,000.

Why? Consider the thought process of a smart person when deciding whether or not to take the course. She thinks "I am smart, so if I take the course, I will certainly pass. Then I will make an extra $60,000 at this job. So my costs are $50,000, and my benefits are $60,000. Sounds like a good deal."

The stupid person, on the other hand, thinks: "As a stupid person, if I take the course, I have a 50% chance of passing and making $60,000 extra, and a 50% chance of failing and making $0 extra. My expected benefit is $30,000, but my expected cost is $50,000. I'll stay out of school and take the $40,000 salary for non-graduates."

…assuming that stupid people all know they're stupid, and that they're all perfectly rational experts at game theory, to name two of several dubious premises here. Yet despite its flaws, this model does give some interesting results. For example, it suggests that rational employers will base decisions upon - and rational employees enroll in - college courses, even if those courses teach nothing of any value. So an investment bank might reject someone who had no college education, even while hiring someone who studied Art History, not known for its relevance to derivative trading.

We'll return to the specific example of education later, but for now it is more important to focus on the general definition that X signals Y if X is more likely to be true when Y is true than when Y is false. Amoral self-interested agents after the $60,000 salary bonus for intelligence, whether they are smart or stupid, will always say "Yes, I'm smart" if you ask them. So saying "I am smart" is not a signal of intelligence. Having a college degree is a signal of intelligence, because a smart person is more likely to get one than a stupid person.

Life frequently throws us into situations where we want to convince other people of something. If we are employees, we want to convince bosses we are skillful, honest, and hard-working. If we run the company, we want to convince customers we have superior products. If we are on the dating scene, we want to show potential mates that we are charming, funny, wealthy, interesting, you name it.

In some of these cases, mere assertion goes a long way. If I tell my employer at a job interview that I speak fluent Spanish, I'll probably get asked to talk to a Spanish-speaker at my job, will either succeed or fail, and if I fail will have a lot of questions to answer and probably get fired - or at the very least be in more trouble than if I'd just admitted I didn't speak Spanish to

begin with. Here society and its system of reputational penalties help turn mere assertion into a credible signal: asserting I speak Spanish is costlier if I don't speak Spanish than if I do, and so is believable.

In other cases, mere assertion doesn't work. If I'm at a seedy bar looking for a one-night stand, I can tell a girl I'm totally a multimillionaire and feel relatively sure I won't be found out until after that one night - and so in this she would be naive to believe me, unless I did something only a real multimillionaire could, like give her an expensive diamond necklace.

How expensive a diamond necklace, exactly? To absolutely prove I am a millionaire, only a million dollars worth of diamonds will do; $10,000 worth of diamonds could in theory come from anyone with at least $10,000. But in practice, people only care so much about impressing a girl at a seedy bar; if everyone cares about the same amount, the amount they'll spend on the signal depends mostly on their marginal utility of money, which in turn depends mostly on how much they have. Both a millionaire and a tenthousandaire can afford to buy $10,000 worth of diamonds, but only the millionaire can afford to buy $10,000 worth of diamonds on a whim. If in general people are only willing to spend 1% of their money on an impulse gift, then $10,000 is sufficient evidence that I am a millionaire.

But when the stakes are high, signals can get prohibitively costly. If a dozen millionaires are wooing Helen of Troy, the most beautiful woman in the world, and willing to spend arbitrarily much money on her - and if they all believe Helen will choose the richest among them - then if I only spend $10,000 on her I'll be outshone by a millionaire who spends the full million. Thus, if I want any chance with her at all, then

even if I am genuinely the richest man around I might have to squander my entire fortune on diamonds.

This raises an important point: *signaling can be really horrible*. What if none of us are entirely sure how much Helen's other suitors have? It might be rational for all of us to spend everything we have on diamonds for her. Then twelve millionaires lose their fortunes, eleven of them for nothing. And this isn't some kind of wealth transfer - for all we know, Helen might not even like diamonds; maybe she locks them in her jewelry box after the wedding and never thinks about them again. It's about as economically productive as digging a big hole and throwing money into it.

If all twelve millionaires could get together beforehand and compare their wealth, and agree that only the wealthiest one would woo Helen, then they could all save their fortunes and the result would be exactly the same: Helen marries the wealthiest. If all twelve millionaires are remarkably trustworthy, maybe they can pull it off. But if any of them believe the others might lie about their wealth, or that one of the poorer men might covertly break their pact and woo Helen with gifts, then they've got to go through with the whole awful "everyone wastes everything they have on shiny rocks" ordeal.

Examples of destructive signaling are not limited to hypotheticals. Even if one does not believe Jared Diamond's hypothesis that Easter Island civilization collapsed after chieftains expended all of their resources trying to out-signal each other by building larger and larger stone heads, one can look at Nikolai Roussanov's study on how the dynamics of signaling games in US minority communities encourage conspicuous consumption and prevent members of those communities from investing in education and other important goods.

The Art of Strategy even advances the surprising hypothesis that corporate advertising can be a form of signaling. When a company advertises during the Super Bowl or some other high-visibility event, it costs a lot of money. To be able to afford the commercial, the company must be pretty wealthy; which in turn means it probably sells popular products and isn't going to collapse and leave its customers in the lurch. And to want to afford the commercial, the company must be pretty confident in its product: advertising that you should shop at Wal-Mart is more profitable if you shop at Wal-Mart, love it, and keep coming back than if you're likely to go to Wal-Mart, hate it, and leave without buying anything. This signaling, too, can become destructive: if every other company in your industry is buying Super Bowl commercials, then none of them have a comparative advantage and they're in exactly the same relative position as if none of them bought Super Bowl commercials - throwing money away just as in the diamond example.

Most of us cannot afford a Super Bowl commercial or a diamond necklace, and less people may build giant stone heads than during Easter Island's golden age, but a surprising amount of everyday life can be explained by signaling. For example, why did about 50% of readers get a mental flinch and an overpowering urge to correct me when I used "less" instead of "fewer" in the sentence above? According to Paul Fussell's "Guide Through The American Class System" (ht SIAI mailing list), nitpicky attention to good grammar, even when a sentence is perfectly clear without it, can be a way to signal education, and hence intelligence and probably social class. I would not dare to summarize Fussell's guide here, but it shattered my illusion that I mostly avoid thinking about class signals, and instead convinced me that pretty much everything

I do from waking up in the morning to going to bed at night is a class signal. On flowers:

> *Anyone imagining that just any sort of flowers can be presented in the front of a house without status jeopardy would be wrong. Upper-middle-class flowers are rhododendrons, tiger lilies, amaryllis, columbine, clematis, and roses, except for bright-red ones. One way to learn which flowers are vulgar is to notice the varieties favored on Sunday-morning TV religious programs like Rex Humbard's or Robert Schuller's. There you will see primarily geraniums (red are lower than pink), poinsettias, and chrysanthemums, and you will know instantly, without even attending to the quality of the discourse, that you are looking at a high-prole setup. Other prole flowers include anything too vividly red, like red tulips. Declassed also are phlox, zinnias, salvia, gladioli, begonias, dahlias, fuchsias, and petunias. Members of the middle class will sometimes hope to mitigate the vulgarity of bright-red flowers by planting them in a rotting wheelbarrow or rowboat displayed on the front lawn, but seldom with success.*

Seriously, read the essay.

In conclusion, a signal is a method of conveying information among not-necessarily-trustworthy parties by performing an action which is more likely or less costly if the information is true than if it is not true. Because signals are often costly, they can sometimes lead to a depressing waste of resources, but in other cases they may be the only way to believably convey important information.

Bargaining and Auctions

Some people have things. Other people want them. Economists agree that the eventual price will be set by supply and demand, but both parties have tragically misplaced their copies of the *Big Book Of Levels Of Supply And Demand For All Goods*. They're going to have to decide on a price by themselves.

When the transaction can be modeled by the interaction of one seller and one buyer, this kind of decision usually looks like bargaining. When it's best modeled as one seller and multiple buyers (or vice versa), the decision usually looks like an auction. Many buyers and many sellers produce a marketplace, but this is complicated and we'll stick to bargains and auctions for now.

Simple bargains bear some similarity to the Ultimatum Game. Suppose an antique dealer has a table she values at $50, and I go to the antique store and fall in love with it, believing it will add $400 worth of classiness to my room. The dealer should never sell for less than $50, and I should never buy for more than $400, but any value in between would benefit both of us. More specifically, it would give us a combined $350 profit. The remaining question is how to divide that $350 pot.

If I make an offer to buy at $60, I'm proposing to split the pot "$10 for you, $340 for me". If the dealer makes a counter-offer of $225, she's offering "$175 for you, $175 for me" - or an even split.

Each round of bargaining resembles the Ultimatum Game because one player proposes to split a pot, and the other player accepts or rejects. If the other player rejects the offer (for

example, the dealer refuses to sell it for $60) then the deal falls through and neither of us gets any money.

But bargaining is unlike the Ultimatum Game for several reasons. First, neither player is the designated "offer-maker"; either player may begin by making an offer. Second, the game doesn't end after one round; if the dealer rejects my offer, she can make a counter-offer of her own. Third, and maybe most important, neither player is exactly sure about the size of the pot: I don't walk in knowing that the dealer bought the table for $50, and I may not really be sure I value the table at $400.

Our intuition tells us that the fairest method is to split the profits evenly at a price of $225. This number forms a useful Schelling point (remember those?) that prevents the hassle of further bargaining.

The *Art of Strategy* (see the beginning of Ch. 11) includes a proof that an even split is the rational choice under certain artificial assumptions. Imagine a store selling souvenirs for the 2012 Olympics. They make $1000/day each of the sixteen days the Olympics are going on. Unfortunately, the day before the Olympics, the workers decide to strike; the store will make no money without workers, and they don't have enough time to hire scabs.

Suppose Britain has some very strange labor laws that mandate the following negotiation procedure: on each odd numbered day of the Olympics, the labor union representative will approach the boss and make an offer; the boss can either accept it or reject it. On each even numbered day, the boss makes the offer to the labor union.

So if the negotiations were to drag on to the sixteenth and last day of the Olympics, on that even-numbered day the boss would approach the labor union rep. They're both the sort of

straw man rationalists who would take 99-1 splits on the Ultimatum Game, so she offers the labor union rep $1 of the $1000. Since it's the last day of the Olympics and she's a straw man rationalist, the rep accepts.

But on the fifteenth day of the Olympics, the labor union rep will approach the boss. She knows that if no deal is struck today, she'll end out with $1 and the boss will end out with $999. She has to convince the boss to accept a deal on the fifteenth day instead of waiting until the sixteenth. So she offers $1 of the profits from the fifteenth day to the boss, with the labor union keeping the rest; now their totals are $1000 for the workers, $1000 for the boss. Since $1000 is better than $999, the boss agrees to these terms and the strike is ended on the fifteenth day.

We can see by this logic that on odd numbered days the boss and workers get the same amount, and on even numbered days the boss gets more than the workers but the ratio converges to 1:1 as the length of the negotiations increase. If they were negotiating an indefinite contract, then even if the boss made the first move we might expect her to offer an even split.

So both some intuitive and some mathematical arguments lead us to converge on this idea of an even split of the sort that gives us the table for $225. But if I want to be a "hard bargainer" - the kind of person who manages to get the table for less than $225 - I have a couple of things I could try.

I could deceive the seller as to how much I valued the table. This is a pretty traditional bargaining tactic: "That old piece of junk? I'd be doing you a favor for taking it off your hands." Here I'm implicitly claiming that the dealer must have paid less than $50, and that I would get less than $400 worth of value. If the dealer paid $20 and I'd only value it to the tune of

$300, then splitting the profit evenly would mean a final price of $160. The dealer could then be expected to counter my move with his own claim as to the table's value: "$160? Do I look like I was born yesterday? This table was old in the time of the Norman Conquest! Its wood comes from a tree that grows on an enchanted island in the Freptane Sea which appears for only one day every seven years!" The final price might be determined by how plausible we each considered the other's claims.

Or I could rig the Ultimatum Game. Used car dealerships are notorious for adding on "extras" after you've agreed on a price over the phone ("Well yes, we agreed the car was $5999, but if you want a steering wheel, that costs another $200.") Somebody (possibly an LWer?) proposed showing up to the car dealership without any cash or credit cards, just a check made out for the agreed-upon amount; the dealer now has no choice but to either take the money or forget about the whole deal. In theory, I could go to the antique dealer with a check made out for $60 and he wouldn't have a lot of options (though do remember that people usually reject ultimata of below about 70-30). The classic bargaining tactic of "I am but a poor chimney sweep with only a few dollars to my name and seven small children to feed and I could never afford a price above $60" seems closely related to this strategy.

And although we're still technically talking about transactions with only one buyer and seller, the mere threat of another seller can change the balance of power drastically. Suppose I tell the dealer I know of another dealer who sells modern art for a fixed price of $300, and that the modern art would add exactly as much classiness to my room as this antique table - that is, I only want one of the two and I'm indifferent between them. Now we're no longer talking about coming up with a

price between $50 and $400 - anything over $300 and I'll reject it and go to the other guy. Now we're talking about splitting the $250 profit between $50 and $300, and if we split it evenly I should expect to pay $175.

(why not $299? After all, the dealer knows $299 is better than my other offer. Because we're still playing the Ultimatum Game, that's why. And if it was $299, then having a second option - art that I like as much as the table - would actually make my bargaining position worse - after all, I was getting it for $225 before.)

Negotiation gurus call this backup option the BATNA ("Best Alternative To Negotiated Agreement") and consider it a useful thing to have. If only one participant in the negotiation has a BATNA greater than zero, that person is less desperate, needs the agreement less, and can hold out for a better deal - just as my $300 art allowed me to lower the asking price of the table from $225 to $175.

This "one buyer, one seller" model is artificial, but from here we can start to see how the real world existence of other buyers and sellers serve as BATNAs for both parties and how such negotiations eventually create the supply and demand of the marketplace.

The remaining case is one seller and multiple buyers (or vice versa). Here the seller's BATNA is "sell it to the other guy", and so a successful buyer must beat the other guy's price. In practice, this takes the form of an auction (why is this different than the previous example? Partly because in the previous example, we were comparing a negotiable commodity - the table - to a fixed price commodity - the art.)

How much should you bid at an auction? In the so-called English auction (the classic auction where a crazy man stands

at the front shouting “Eighty!!! Eighty!!! We have eighty!!! Do I hear eighty-five?!? Eighty-five?!? Eighty-five to the man in the straw hat!!! Do I hear ninety?!?) the answer should be pretty obvious: keep bidding infinitesmally more than the last guy until you reach your value for the product, then stop. For example, with the $400 table, keep bidding until the price approaches $400.

But what about a sealed-bid auction, where everyone hands the auctioneer their bid and the auctioneer gives the product to the highest? Or what about the so-called “Dutch auction” where the auctioneer starts high and goes lower until someone bites (“A hundred?!? Anyone for a hundred?!? No?!? Ninety-five?!? Anyone for…yes?!? Sold for ninety-five to the man in the straw hat!!!).

The rookie mistake is to bid the amount you value the product. Remember, economists define “the amount you value the product” as “the price at which you would be indifferent between having the product and just keeping the money”. If you go to an auction planning to bid your true value, you should expect to get absolutely zero benefit out of the experience. Instead, you should bid infinitesimally more than what you predict the next highest bidder will pay, as long as this is below your value.

Thus, the auction beloved by economists as perhaps the purest example of auction forms is the Vickrey, in which everyone submits a sealed bid, the highest bidder wins, and she pays the amount of the second-highest bid. This auction has a certain very elegant property, which is that here the dominant strategy is to bid your true value. Why?

Suppose you value a table at $400. If you try to game the system by bidding $350 instead of $400, you may lose out

and can at best break even. Why? Because if the highest other bid was above $400, you wouldn't win the table in either case, and your ploy profits you nothing. And if the highest other bid was between $350 and $400 (let's say $375), now you lose the table and make $0 profit, as opposed to the $25 profit you would have made if you had bid your true value of $400, won, and paid the second-highest bid of $375. And if everyone else is below $350 (let's say $300) then you would have paid $300 in either case, and again your ploy profits you nothing. Bid above your true valuation (let's say $450) and you face similar consequences: either you wouldn't have gotten the table anyway, you get the table for the same amount as before, or you get the table for a value between $400 and $450 and now you're taking a loss.

In the real world, English, Dutch, sealed-bid and Vickrey auctions all differ a little in ways like how much information they give the bidders about each other, or whether people get caught up in the excitement of bidding, or what to do when you don't really know your true valuation. But in simplified rational models, they all end at an identical price: the true valuation of the second-highest bidder.

In conclusion, the gentlemanly way to bargain is to split the difference in profits between your and your partner's best alternative to an agreement, and gentlemanly auctions tend to end at the value of the second-highest participant. Some less gentlemanly alternatives are also available and will be discussed later.

Imperfect Voting Systems

Stalin once (supposedly) said that “He who casts the votes determines nothing; he who counts the votes determines everything “ But he was being insufficiently cynical. He who chooses the voting system may determine just as much as the other two players.

The Art of Strategy gives some good examples of this principle: here’s an adaptation of one of them. Three managers are debating whether to give a Distinguished Employee Award to a certain worker. If the worker gets the award, she must receive one of two prizes: a $50 gift certificate, or a $10,000 bonus.

One manager loves the employee and wants her to get the $10,000; if she can’t get the $10,000, she should at least get a gift certificate. A second manager acknowledges her contribution but is mostly driven by cost-cutting; she’d be happiest giving her the gift certificate, but would rather refuse to recognize her entirely than lose $10,000. And the third manager dislikes her and doesn’t want to recognize her at all - but she also doesn’t want the company to gain a reputation for stinginess, so if she gets recognized she’d rather give her the $10,000 than be so pathetic as to give her the cheap certificate.

The managers arrange a meeting to determine the employee’s fate. If the agenda tells them to vote for or against giving her an award, and then proceed to determine the prize afterwards if she wins, then things will not go well for the employee. Why not? Because the managers reason as follows: if she gets the award, Manager 1 and Manager 3 will vote for the $10,000 prize, and Manager 2 will vote for the certificate. Therefore, voting for her to get the award is practically the same as voting

for her to get the $10,000 prize. That means Manager 1, who wants her to get the prize, will vote yes on the award, but Managers 2 and 3, who both prefer no award to the $10,000, will strategically vote not to give her the award. Result: she doesn't get recognized for her distinguished service.

But suppose the employee involved happens to be the secretary arranging the meeting where the vote will take place. She makes a seemingly trivial change to the agenda: the managers will vote for what the prize should be first, and then vote on whether to give it to her.

If the managers decide the appropriate prize is $10,000, then the motion to give the award will fail for exactly the same reasons it did above. But if the managers decide the certificate is appropriate, then Manager 1 and 2, who both prefer the certificate to nothing, will vote in favor of giving the award. So the three managers, thinking strategically, realize that the decision before them, which looks like "$10 grand or certificate", is really "No award or certificate". Since 1 and 2 both prefer the certificate to nothing, they vote that the certificate is the appropriate prize (even though Manager 1 doesn't really believe this) and the employee ends out with the gift certificate.

But if the secretary is really smart, she may set the agenda as follows: The managers first vote whether or not to give $10,000, and if that fails, they next vote whether or not to give the certificate; if both votes fail the employee gets nothing. Here the managers realize that if the first vote (for $10,000) fails, the next vote (certificate or nothing) will pass, since two managers prefer certificate to nothing as mentioned before. So the true choice in the first vote is "$10,000 versus certificate". Since two managers (1 and 3) prefer the $10,000 to the

certificate, those two start by voting to give the full $10,000, and this is what the employee gets.

So we see that all three options are possible outcomes, and that the true power rests not in the hands of any individual manager, but in the secretary who determines how the voting takes place.

Americans have a head start in understanding the pitfalls of voting systems thanks to the so-called two party system. Every four years, they face quandaries like “If leftists like me vote for Nader instead of Gore just because we like him better, are we going to end up electing Bush because we’ve split the leftist vote?”

Empirically, yes. The 60,000 Florida citizens who voted Green in 2000 didn’t elect Nader. However, they did make Gore lose to Bush by a mere 500 votes. The last post discussed a Vickrey auction, a style of auction in which you have have no incentive to bid anything except your true value. Wouldn’t it be nice if we had an electoral system with the same property: one where you should always vote for the candidate you actually support? If such a system existed, we would have ample reason to institute it and could rest assured that no modern-day Stalin was manipulating us via the choice of voting system we used.

Some countries do claim to have better systems than the simple winner-takes-all approach of the United States. My own adopted homeland of Ireland uses a system called “single transferable vote” (also called instant-runoff vote), in which voters rank the X candidates from 1 to X. If a candidate has the majority of first preference votes (or a number of first preference votes greater than the number of positions to fill divided by the number of candidates, in elections with multiple potential winners like legislative elections), then that

candidate wins and any surplus votes go to their voters' next preference. If no one meets the quota, then the least popular candidate is eliminated and their second preference votes become first preferences. The system continues until all available seats are full.

For example, suppose I voted (1: Nader), (2: Gore), (3: Bush). The election officials tally all the votes and find that Gore has 49 million first preferences, Bush has 50 million, and Nader has 5 million. There's only one presidency, so a candidate would have to have a majority of votes (greater than 52 million out of 104 million) to win. Since no one meets that quota, the lowest ranked candidate gets eliminated - in this case, Nader. My vote now goes to my second preference, Gore. If 4 million Nader voters put Gore second versus 1 million who put Bush second, the tally's now at 53 million Gore, 51 million Bush. Gore has greater than 52 million and wins the election - the opposite result from if we'd elected a president the traditional way.

Another system called Condorcet voting also uses a list of all candidates ranked in order, but uses the information to run mock runoffs between each of them. So a Condorcet system would use the ballots to run a Gore/Nader match (which Gore would win), a Gore/Bush match (which Gore would win), and a Bush/Nader match (which Bush would win). Since Gore won all of his matches, he becomes President. This becomes complicated when no candidate wins all of his matches (imagine Gore beating Nader, Bush beating Gore, but Nader beating Bush in a sort of Presidential rock-paper-scissors.) Condorcet voting has various options to resolve this; some systems give victory to the candidate whose greatest loss was by the smallest margin, and others to candidates who defeated the greatest number of other candidates.

Do these systems avoid the strategic voting that plagues American elections? No. For example, both Single Transferable Vote and Condorcet voting sometimes provide incentives to rank a candidate with a greater chance of winning higher than a candidate you prefer - that is, the same "vote Gore instead of Nader" dilemma you get in traditional first-past-the-post.

There are many other electoral systems in use around the world, including several more with ranking of candidates, a few that do different sorts of runoffs, and even some that ask you to give a numerical rating to each candidate (for example "Nader 10, Gore 6, Bush -100000"). Some of them even manage to eliminate the temptation to rank a non-preferred candidate first. But these work only at the expense of incentivizing other strategic manuevers, like defining "approved candidate" differently or exaggerating the difference between two candidates.

So is there any voting system that automatically reflects the will of the populace in every way without encouraging tactical voting? No. Various proofs, including the Gibbard-Satterthwaite Theorem and the better-known Arrow Impossibility Theorem show that many of the criteria by which we would naturally judge voting systems are mutually incompatible and that all reasonable systems must contain at least some small element of tactics (one example of an unreasonable system that eliminates tactical voting is picking one ballot at random and determining the results based solely on its preferences; the precise text of the theorem rules out "nondeterministic or dictatorial" methods).

This means that each voting system has its own benefits and drawbacks, and that which one people use is largely a matter of preference. Some of these preferences reflect genuine

concern about the differences between voting systems: for example, is it better to make sure your system always elects the Condorcet winner, even if that means the system penalizes candidates who are too similar to other candidates? Is it better to have a system where you can guarantee that participating in the election always makes your candidate more likely to win, or one where you can be sure that everyone voting exactly the opposite will never elect the same candidate?

But in practice, these preferences tend to be political and self-interested. This was recently apparent in Britain, which voted last year on a referendum to change the voting system. The Liberal Democrats, who were perpetually stuck in the same third-place situation as Nader in the States, supported a change to a form of instant runoff voting which would have made voting Lib Dem a much more palatable option; the two major parties opposed it probably for exactly that reason.

Although no single voting system is mathematically perfect, several do seem to do better on the criteria that real people care about; look over Wikipedia's section on the strengths and weaknesses of different voting systems to see which one looks best.

Game Theory as a Dark Art

One of the most charming features of game theory is the almost limitless depths of evil to which it can sink.

Your garden-variety evils act against your values. Your better class of evil, like Voldemort and the folk-tale version of Satan, use your greed to trick *you* into acting against *your own* values, then grab away the promised reward at the last moment. But even demons and dark wizards can only do this once or twice before most victims wise up and decide that taking their advice is a bad idea. Game theory can force you to betray your deepest principles for no lasting benefit again and again, and still leave you convinced that your behavior was rational.

Some of the examples in this post probably wouldn't work in reality; they're more of a *reductio ad absurdum* of the so-called *homo economicus* who acts free from any feelings of altruism or trust. But others are lifted directly from real life where seemingly intelligent people genuinely fall for them. And even the ones that don't work with real people might be valuable in modeling institutions or governments.

Of the following examples, the first three are from *The Art of Strategy*; the second three are relatively classic problems taken from around the Internet. A few have been mentioned in the comments here already and are reposted for people who didn't catch them the first time.

The Evil Plutocrat

You are an evil plutocrat who wants to get your pet bill - let's say a law that makes evil plutocrats tax-exempt - through the US Congress. Your usual strategy would be to bribe the Congressmen involved, but that would be pretty costly - Congressmen no longer come cheap. Assume all Congressmen act in their own financial self-interest, but that absent any financial self-interest they will grudgingly default to honestly representing their constituents, who hate your bill (and you personally). Is there any way to ensure Congress passes your bill, without spending any money on bribes at all?

Yes. Simply tell all Congressmen that *if* your bill fails, you will donate some stupendous amount of money to whichever party gave the greatest percent of their votes in favor.

Suppose the Democrats try to coordinate among themselves. They say "If we all oppose the bill, then if even one Republican supports the bill, the Republicans will get lots of money they can spend on campaigning against us. If only one of us supports the bill, the Republicans may anticipate this strategy and two of them may support it. The only way to ensure the Republicans don't gain a massive windfall and wipe the floor with us next election is for most of us to vote for the bill."

Meanwhile, in their meeting, the Republicans think the same thing. The vote ends with most members of Congress supporting your bill, and you don't end up having to pay any money at all.

The Hostile Takeover

You are a ruthless businessman who wants to take over a competitor. The competitor's stock costs $100 a share, and there are 1000 shares, distributed among a hundred investors who each own ten. That means the company ought to cost

$100,000, but you don't have $100,000. You only have $98,000. Worse, another competitor with $101,000 has made an offer for greater than the value of the company: they will pay $101 per share if they end up getting all of the shares. Can you still manage to take over the company?

Yes. You can make what is called a two-tiered offer. Suppose all investors get a chance to sell shares simultaneously. You will pay $105 for 500 shares - better than they could get from your competitor - but only pay $90 for the other 500. If you get fewer than 500 shares, all will sell for $105; if you get more than 500, you will start by distributing the $105 shares evenly among all investors who sold to you, and then distribute out as many of the $90 shares as necessary (leaving some $90 shares behind except when all investors sell to you) . And you will do this whether or not you succeed in taking over the company - if only one person sells you her share, then that one person gets $105.

Suppose an investor believes you're not going to succeed in taking over the company. That means you're not going to get over 50% of shares. That means the offer to buy 500 shares for $105 will still be open. That means the investor can either sell her share to you (for $105) or to your competitor (for $101). Clearly, it's in this investor's self-interest to sell to you.

Suppose the investor believes you will succeed in taking over the company. That means your competitor will not take over the company, and its $101 offer will not apply. That means that the new value of the shares will be $90, the offer you've made for the second half of shares. So they will get $90 if they don't sell to you. How much will they get if they do sell to you? They can expect half of their ten shares to go for $105 and half to go for $90; they will get a total of $97.50 per share. $97.50 is better than $90, so their incentive is to sell to you.

Suppose the investor believes you are right on the cusp of taking over the company, and her decision will determine the outcome. In that case, you have at most 499 shares. When the investor gives you her 10 shares, you will end up with 509 - 500 of which are $105 shares and 9 of which are $90 shares. If these are distributed randomly, investors can expect to make on average $104.73 per share, compared to $101 if your competitor buys the company.

Since all investors are thinking along these lines, they all choose to buy shares from you instead of your competitor. You pay out an average of $97.50 per share, and take over the company for $97,500, leaving $500 to spend on the victory party.

The stockholders, meanwhile, are left wondering why they just all sold shares for $97.50 when there was someone else who was promising them $101.

The Hostile Takeover, Part II

Your next target is a small family-owned corporation that has instituted what they consider to be invincible protection against hostile takeovers. All decisions are made by the Board of Directors, who serve for life. Although shareholders vote in the new members of the Board after one of them dies or retires, Board members can hang on for decades. And all decisions about the Board, impeachment of its members, and enforcement of its bylaws are made by the Board itself, with members voting from newest to most senior.

So you go about buying up 51% of the stock in the company, and sure enough, a Board member retires and is replaced by one of your lackeys. This lackey can propose procedural changes to the Board, but they have to be approved by majority vote. And at the moment the other four directors hate

you with a vengeance, and anything you propose is likely to be defeated 4-1. You need those other four windbags out of there, and soon, but they're all young and healthy and unlikely to retire of their own accord.

The obvious next step is to start looking for a good assassin. But if you can't find one, is there any way you can propose mass forced retirement to the Board and get them to approve it by majority vote? Even better, is there any way you can get them to approve it unanimously, as a big "f#@& you" to whoever made up this stupid system?

Yes. Your lackey proposes as follows: "I move that we vote upon the following: that if this motion passes unanimously, all members of the of the Board resign immediately and are given a reasonable compensation; that if this motion passes 4-1 that the Director who voted against it must retire without compensation, and the four directors who voted in favor may stay on the Board; and that if the motion passes 3-2, then the two 'no' voters get no compensation and the three 'yes' voters may remain on the board and will also get a spectacular prize - to wit, our company's 51% share in your company divided up evenly among them."

Your lackey then votes "yes". The second newest director uses backward reasoning as follows:

Suppose that the vote were tied 2-2. The most senior director would prefer to vote "yes", because then she gets to stay on the Board and gets a bunch of free stocks.

But knowing that, the second most senior director (SMSD) will also vote 'yes'. After all, when the issue reaches the SMSD, there will be one of the following cases:

1. If there is only one yes vote (your lackey's), the SMSD stands to gain from voting yes, knowing that will produce a 2-

2 tie and make the most senior director vote yes to get her spectacular compensation. This means the motion will pass 3-2, and the SMSD will also remain on the board and get spectacular compensation if she votes yes, compared to a best case scenario of remaining on the board if she votes no.

2. If there are two yes votes, the SMSD must vote yes - otherwise, it will go 2-2 to the most senior director, who will vote yes, the motion will pass 3-2, and the SMSD will be forced to retire without compensation.

3. And if there are three yes votes, then the motion has already passed, and in all cases where the second most senior director votes “no”, she is forced to retire without compensation. Therefore, the second most senior director will always vote “yes”.

Since your lackey, the most senior director, and the second most senior director will always vote “yes”, we can see that the other two directors, knowing the motion will pass, must vote “yes” as well in order to get any compensation at all. Therefore, the motion passes unanimously and you take over the company at minimal cost.

The Dollar Auction

You are an economics professor who forgot to go to the ATM before leaving for work, and who has only $20 in your pocket. You have a lunch meeting at a very expensive French restaurant, but you’re stuck teaching classes until lunchtime and have no way to get money. Can you trick your students into giving you enough money for lunch in exchange for your $20, without lying to them in any way?

Yes. You can use what’s called an all-pay auction, in which several people bid for an item, as in a traditional auction, but

everyone pays their bid regardless of whether they win or lose (in a common variant, only the top two bidders pay their bids).

Suppose one student, Alice, bids $1. This seems reasonable - paying $1 to win $20 is a pretty good deal. A second student, Bob, bids $2. Still a good deal if you can get a twenty for a tenth that amount.

The bidding keeps going higher, spurred on by the knowledge that getting a $20 for a bid of less than $20 would be pretty cool. At some point, maybe Alice has bid $18 and Bob has bid $19.

Alice thinks: “What if I raise my bid to $20? Then certainly I would win, since Bob would not pay more than $20 to get $20, but I would only break even. However, breaking even is better than what I’m doing now, since if I stay where I am Bob wins the auction and I pay $18 without getting anything.” Therefore Alice bids $20.

Bob thinks “Well, it sounds pretty silly to bid $21 for a twenty dollar bill. But if I do that and win, I only lose a dollar, as opposed to bowing out now and losing my $19 bid.” So Bob bids $21.

Alice thinks “If I give up now, I’ll lose a whole dollar. I know it seems stupid to keep going, but surely Bob has the same intuition and he’ll give up soon. So I’ll bid $22 and just lose two dollars…”

It’s easy to see that the bidding could in theory go up with no limits but the players’ funds, but in practice it rarely goes above $200.

…yes, $200. Economist Max Bazerman claims that of about 180 such auctions, seven have made him more than $100 (ie

$50 from both players) and his highest take was $407 (ie over $200 from both players).

In any case, you're probably set for lunch. If you're not, take another $20 from your earnings and try again until you are - the auction gains even more money from people who have seen it before than it does from naive bidders (!) Bazerman, for his part, says he's made a total of $17,000 from the exercise.

At that point you're starting to wonder why no one has tried to build a corporation around this, and unsurprisingly, the online auction site Swoopo appears to be exactly that. More surprisingly, they seem to have gone bankrupt last year, suggesting that maybe H.L. Mencken was wrong and someone *has* gone broke underestimating people's intelligence.

The Bloodthirsty Pirates

You are a pirate captain who has just stolen $17,000, denominated entirely in $20 bills, from a very smug-looking game theorist. By the Pirate Code, you as the captain may choose how the treasure gets distributed among your men. But your first mate, second mate, third mate, and fourth mate all want a share of the treasure, and demand on threat of mutiny the right to approve or reject any distribution you choose.You expect they'll reject anything too lopsided in your favor, which is too bad, because that was totally what you were planning on.

You remember one fact that might help you - your crew, being bloodthirsty pirates, all hate each other and actively want one another dead. Unfortunately, their greed seems to have overcome their bloodlust for the moment, and as long as there are advantages to coordinating with one another, you won't be

able to turn them against their fellow sailors. Doubly unfortunately, they also actively want you dead.

You think quick. “Aye,” you tell your men with a scowl that could turn blood to ice, “ye can have yer votin’ system, ye scurvy dogs” (you’re that kind of pirate). “But here’s the rules: I propose a distribution. Then you all vote on whether or not to take it. If a majority of you, or even half of you, vote ‘yes’, then that’s how we distribute the treasure. But if you vote ‘no’, then I walk the plank to punish me for my presumption, and the first mate is the new captain. He proposes a new distribution, and again you vote on it, and if you accept then that’s final, and if you reject it he walks the plank and the second mate becomes the new captain. And so on.”

Your four mates agree to this proposal. What distribution should you propose? Will it be enough to ensure your comfortable retirement in Jamaica full of rum and wenches?

Yes. Surprisingly, you can get away with proposing that you get $16,960, your first mate gets nothing, your second mate gets $20, your third mate gets nothing, and your fourth mate gets $20 - and you will still win 3 -2.

The fourth mate uses backward reasoning like so: Suppose there were only two pirates left, me and the third mate. The third mate wouldn’t have to promise me anything, because if he proposed all $17,000 for himself and none for me, the vote would be 1-1 and according to the original rules a tie passes. Therefore this is a better deal than I would get if it were just me and the third mate.

But suppose there were three pirates left, me, the third mate, and the second mate. Then the second mate would be the new captain, and he could propose $16,980 for himself, $0 for the third mate, and $20 for me. If I vote no, then it reduces to the

previous case in which I get nothing. Therefore, I should vote yes and get $20. Therefore, the final vote is 2-1 in favor.

But suppose there were four pirates left: me, the third mate, the second mate, and the first mate. Then the first mate would be the new captain, and he could propose $16,980 for himself, $20 for the third mate, $0 for the second mate, and $0 for me. The third mate knows that if he votes no, this reduces to the previous case, in which he gets nothing. Therefore, he should vote yes and get $20. Therefore, the final vote is 2-2, and ties pass.

(He might also propose $16980 for himself, $0 for the second mate, $0 for the third mate, and $20 for me. But since he knows I am a bloodthirsty pirate who all else being equal wants him dead, I would vote no since I could get a similar deal from the third mate and make the first mate walk the plank in the bargain. Therefore, he would offer the $20 to the third mate.)

But in fact there are five pirates left: me, the third mate, the second mate, the first mate, and the captain. The captain has proposed $16,960 for himself, $20 for the second mate, and $20 for me. If I vote no, this reduces to the previous case, in which I get nothing. Therefore, I should vote yes and get $20.

(The captain would avoid giving the $20s to the third and fourth rather than to the second and fourth mates for a similar reason to the one given in the previous example - all else being equal, the pirates would prefer to watch him die.)

The second mate thinks along the same lines and realizes that if he votes no, this reduces to the case with the first mate, in which the second mate also gets nothing. Therefore, he too votes yes.

Since you, as the captain, obviously vote yes as well, the distribution passes 3-2. You end up with $16,980, and your crew, who were so certain of their ability to threaten you into sharing the treasure, each end up with either a single $20 or nothing.

The Prisoners' Dilemma, Redux

This sequence previously mentioned the popularity of Prisoners' Dilemmas as gimmicks on TV game shows. In one program, Golden Balls, contestants do various tasks that add money to a central "pot". By the end of the game, only two contestants are left, and are offered a Prisoners' Dilemma situation to split the pot between them. If both players choose to "Split", the pot is divided 50-50. If one player "Splits" and the other player "Steals", the stealer gets the entire pot. If both players choose to "Steal", then no one gets anything. The two players are allowed to talk to each other before making a decision, but like all Prisoner's Dilemmas, the final choice is made simultaneously and in secret.

You are a contestant on this show. You are actually not all that evil - you would prefer to split the pot rather than to steal all of it for yourself - but you certainly don't want to trust the other guy to have the same preference. In fact, the other guy looks a bit greedy. You would prefer to be able to rely on the other guy's rational self-interest rather than on his altruism. Is there any tactic you can use before the choice, when you're allowed to communicate freely, in order to make it rational for him to cooperate?

Yes. In one episode of Golden Balls, a player named Nick successfully meta-games the game by transforming it from the Prisoner's Dilemma (where defection is rational) to the Ultimatum Game (where cooperation is rational)

Nick tells his opponent: “I am going to choose ‘Steal’ on this round.” (He then immediately pressed his button; although the show hid which button he pressed, he only needed to demonstrate that he had committed and his mind could no longer be changed) “If you also choose ‘Steal’, then for certain neither of us gets any money. If you choose ‘Split’, then I get all the money, but immediately after the game, I will give you half of it. You may not trust me on this, and that’s understandable, but think it through. First, there’s no less reason to think I’m trustworthy than if I had just told you I pressed ‘Split’ to begin with, the way everyone else on this show does. And second, now if there’s any chance whatsoever that I’m trustworthy, then that’s some chance of getting the money - as opposed to the zero chance you have of getting the money if you choose ‘Steal’.”

Nick’s evaluation is correct. His opponent can either press ‘Steal’, with a certainty of getting zero, or press ‘Split’, with a nonzero probability of getting his half of the pot depending on Nick’s trustworthiness.

But this solution is not quite perfect, in that one can imagine Nick’s opponent being very convinced that Nick will cheat him, and deciding he values punishing this defection more than the tiny chance that Nick will play fair. That’s why I was so impressed to see cousin_it propose what I think is an even better solution on the Less Wrong thread on the matter:

> This game has multiple Nash equilibria and cheap talk is allowed, so correlated equilibria are possible. Here’s how you implement a correlated equilibrium if your opponent is smart enough:
>
> “We have two minutes to talk, right? I’m going to ask you to flip a coin (visibly to both of us) at the last possible

> moment, the exact second where we must cease talking. If the coin comes up heads, I promise I'll cooperate, you can just go ahead and claim the whole prize. If the coin comes up tails, I promise I'll defect. Please cooperate in this case, because you have nothing to gain by defecting, and anyway the arrangement is fair, isn't it?"

This sort of clever thinking is, in my opinion, the best that game theory has to offer. It shows that game theory need not be only a tool of evil for classical figures of villainy like bloodthirsty pirate captains or corporate raiders or economists, but can also be used to create trust and ensure cooperation between parties with common interests.

VI. Promises and Principles

Beware Trivial Inconveniences

The Great Firewall of China. A massive system of centralized censorship purging the Chinese version of the Internet of all potentially subversive content. Generally agreed to be a great technical achievement and political success even by the vast majority of people who find it morally abhorrent.

I spent a few days in China. I got around it at the Internet cafe by using a free online proxy. Actual Chinese people have dozens of ways of getting around it with a minimum of technical knowledge or just the ability to read some instructions.

The Chinese government isn't losing any sleep over this (although they also don't lose any sleep over murdering political dissidents, so maybe they're just very sound sleepers). Their theory is that by making it a little inconvenient and time-consuming to view subversive sites, they will discourage casual exploration. No one will bother to circumvent it unless they already seriously distrust the Chinese government and are specifically looking for foreign websites, and these people probably know what the foreign websites are going to say anyway.

Think about this for a second. The human longing for freedom of information is a terrible and wonderful thing. It delineates a pivotal difference between mental emancipation and slavery. It has launched protests, rebellions, and revolutions. Thousands have devoted their lives to it, thousands of others have even died for it. And it can be stopped dead in its tracks by requiring people to search for "how to set up proxy" before viewing their anti-government website.

I was reminded of this recently by Eliezer's Less Wrong Progress Report. He mentioned how surprised he was that so many people were posting so much stuff on Less Wrong, when very few people had ever taken advantage of Overcoming Bias' policy of accepting contributions if you emailed them to a moderator and the moderator approved. Apparently all us folk brimming with ideas for posts didn't want to deal with the aggravation.

Okay, in my case at least it was a bit more than that. There's a sense of going out on a limb and drawing attention to yourself, of arrogantly claiming some sort of equivalence to Robin Hanson and Eliezer Yudkowsky. But it's still interesting that this potential embarrassment and awkwardness was enough to keep the several dozen people who have blogged on here so far from sending that "I have something I'd like to post…" email.

Companies frequently offer "free rebates". For example, an $800 television with a $200 rebate. There are a few reasons companies like rebates, but one is that you'll be attracted to the television because it appears to have a net cost only $600, but then filling out the paperwork to get the rebate is too inconvenient and you won't get around to it. This is basically a free $200 for filling out an annoying form, but companies can predict that customers will continually fail to complete it. This might make some sense if you're a high-powered lawyer or someone else whose time is extremely valuable, but most of us have absolutely no excuse.

One last example: It's become a truism that people spend more when they use credit cards than when they use money. This particular truism happens to be true: in a study by Prelec and Simester[1], auction participants bid twice as much for the same prize when using credit than when using cash. The trivial step

of getting the money and handing it over has a major inhibitory effect on your spending habits.

I don't know of any unifying psychological theory that explains our problem with trivial inconveniences. It seems to have something to do with loss aversion, and with the brain's general use of emotion-based hacks instead of serious cost-benefit analysis. It might be linked to akrasia; for example, you might not have enough willpower to go ahead with the unpleasant action of filling in a rebate form, and your brain may assign it low priority because it's hard to imagine the connection between the action and the reward.

But these trivial inconveniences have major policy implications. Countries like China that want to oppress their citizens are already using "soft" oppression to make it annoyingly difficult to access subversive information. But there are also benefits for governments that want to help their citizens.

"Soft paternalism" means a lot of things to a lot of different people. But one of the most interesting versions is the idea of "opt-out" government policies. For example, it would be nice if everyone put money into a pension scheme. Left to their own devices, many ignorant or lazy people might never get around to starting a pension, and in order to prevent these people's financial ruin, there is strong a moral argument for a government-mandated pension scheme. But there's also a strong libertarian argument against that idea; if someone for reasons of their own doesn't want a pension, or wants a different kind of pension, their status as a free citizen should give them that right.

The "soft paternalist" solution is to have a government-mandated pension scheme, but allow individuals to opt-out of

it after signing the appropriate amount of paperwork. Most people, the theory goes, would remain in the pension scheme, because they understand they're better off with a pension and it was only laziness that prevented them from getting one before. And anyone who actually goes through the trouble of opting out of the government scheme would either be the sort of intelligent person who has a good reason not to want a pension, or else deserve what they get[2].

This also reminds me of Robin's IQ-gated, test-requiring would-have-been-banned store, which would discourage people from certain drugs without making it impossible for the true believers to get their hands on them. I suggest such a store be located way on the outskirts of town accessible only by a potholed road with a single traffic light that changes once per presidential administration, have a surly clerk who speaks heavily accented English, and be open between the hours of two and four on weekdays.

Footnotes

1: See Jonah Lehrer's book *How We Decide.* In fact, do this anyway. It's very good.

2: Note also the clever use of the status quo bias here.

Time and Effort Discounting

Related to: Akrasia, hyperbolic discounting, and picoeconomics

If you're tired of studies where you inevitably get deceived, electric shocked, or tricked into developing a sexual attraction to penny jars, you might want to sign up for Brian Wansink's next experiment. He provided secretaries with a month of unlimited free candy at their workplace. The only catch was that half of them got the candy in a bowl on their desk, and half got it in a bowl six feet away. The deskers ate five candies/day more than the six-footers, which the scientists calculated would correspond to a weight gain of over 10 pounds more per year[1].

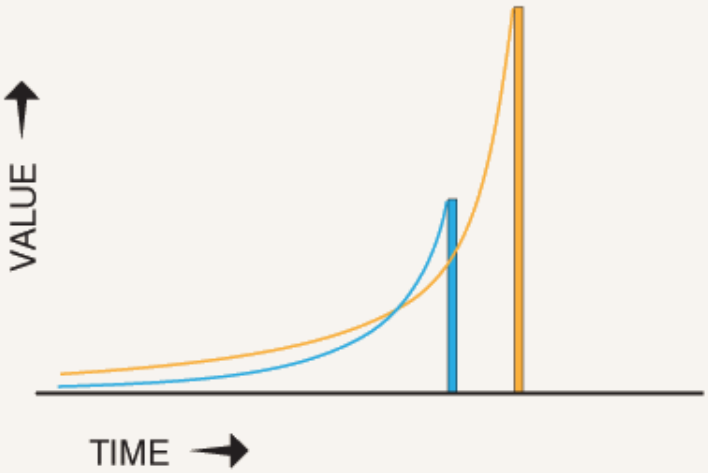
Beware trivial inconveniences (or, in this case, if you don't want to gain weight, beware the lack of them!) Small modifications to the difficulty of obtaining a reward can make big differences in whether the corresponding behavior gets executed.

TIME DISCOUNTING

The best studied example of this is time discounting. When offered two choices, where A will lead to a small reward now and B will lead to a big reward later, people will sometimes choose smaller-sooner rather than larger-later depending on the length of the delay and the size of the difference. For example, in one study, people preferred $250 today to $300 in a year; it took a promise of at least $350 to convince them to wait.

Time discounting was later found to be “hyperbolic”, meaning that the discount amount between two fixed points decreases the further you move those two points into the future. For example, you might prefer $80 today to $100 one week from now, but it’s unlikely you would prefer $80 in one hundred weeks to $100 in one hundred one weeks. Yet this is offering essentially the same choice: wait an extra week for an extra $20. So it’s not enough to say that the discount rate is a constant 20% per week - the discount rate changes depending on what interval of time we’re talking about. If you graph experimentally obtained human discount rates on a curve, they form a hyperbola.

Hyperbolic discounting creates the unpleasant experience of “preference reversals”, in which people can suddenly change their mind on a preference as they move along the hyperbola. For example, if I ask you today whether you would prefer $250 in 2019 or $300 in 2020 (a choice between small reward in 8 years or large reward in 9), you might say the $300 in 2020; if I ask you in 2019 (when it’s a choice between small reward now and large reward in 1 year), you might say no, give me the $250 now. In summary, people prefer larger-later rewards most of the time EXCEPT for a brief period right before they can get the smaller-sooner reward.



Hyperbolic discounting. Notice how the smaller-sooner reward (blue) is considered less valuable than the larger-later reward (orange) EXCEPT just before the smaller-sooner reward can be obtained.

Taken from: George Ainslie, picoeconomics.org

George Ainslie ties this to akrasia and addiction: call the enjoyment of a cigarette in five minutes the smaller-sooner reward, and the enjoyment of not having cancer in thirty years the larger-later reward. You'll prefer to abstain right up until the point where there's a cigarette in front of you and you think "I should smoke this", at which point you will do so.

Discounting can happen on any scale from seconds to decades, and it has previously been mentioned that the second or sub-second level may have disproportionate effects on our actions. Eliezer concentrated on the difficult of changing tasks, but I would add that any task which allows continuous delivery of small amounts of reinforcement with near zero delay can become incredibly addictive even if it isn't all that fun (this is why I usually read all the way through online joke lists, or stay on Reddit for hours). This is also why the XKCD solution to internet addiction - an extension that makes you wait 30 seconds before loading addictive sites - is so useful.

EFFORT DISCOUNTING

Effort discounting is time discounting's lesser-known cousin. It's not obvious that it's an independent entity; it's hard to

disentangle from time discounting (most efforts usually take time) and from garden-variety balancing benefits against costs (most efforts are also slightly costly). There have really been only one or two good studies on it and they don't do much more than say it probably exists and has its own signal in the nucleus accumbens.

Nevertheless, I expect that effort discounting, like time discounting, will be found to be hyperbolic. Many of these trivial inconveniences involve not just time but effort: the secretaries had to actually stand up and walk six feet to get the candy. If a tiny amount of effort held the same power as a tiny amount of time, it would go even further toward explaining garden-variety procrastination.

TIME/EFFORT DISCOUNTING AND UTILITY

Hyperbolic discounting stretches our intuitive notion of "preference" to the breaking point.

Traditionally, discount rates are viewed as just another preference: not only do I prefer to have money, but I prefer to have it now. But hyperbolic discounting shows that we have no single discount rate: instead, we have different preferences for discount rates at different future times.

It gets worse. Time discount rates seem to be different for losses and gains, and different for large amounts vs. small amounts (I gave the example of $250 now being worth $350 in a year, but the same study found that $3000 now is only worth $4000 in a year, and $15 now is worth a whopping $60 in a year). You can even get people to exhibit negative discount rates in certain situations: offer people $10 now, $20 in a month, $30 in two months, and $40 in three months, and they'll prefer it to $40 now, $30 in a month, and so on - maybe because it's nice to think things are only going to get better?

Are there utility functions that can account for this sort of behavior? Of course: you can do a lot of things just by adding enough terms to an equation. But what is the "preference" that the math is describing? When I say I like having money, that seems clear enough: preferring $20 to $15 is not a separate preference than preferring $406 to $405.

But when we discuss time discounting, most of the preferences cited are specific: that I would prefer $100 now to $150 later. Generalizing these preferences, when it's possible at all, takes several complicated equations. Do I really want to discount gains more than losses, if I've never consciously thought about it and I don't consciously endorse it? Sure, there might be such things as unconscious preferences, but saying that the unconscious just loves following these strange equations, in the same way that it loves food or sex or status, seems about as contrived as saying that our robot just really likes switching from blue-minimization to yellow-minimization every time we put a lens on its sensor.

It makes more sense to consider time and effort discounting as describing reward functions and not utility functions. The brain estimates the value of reward in neural currency using these equations (or a neural network that these equations approximate) and then people execute whatever behavior has been assigned the highest reward.

Footnotes

1: Also cited in the same Nutrition Action article: if the candy was in a clear bowl, participants ate on average two/day more than if the candy was in an opaque bowl.

Applied Picoeconomics

Related to: Akrasia, Hyperbolic Discounting, and Picoeconomics, Fix It And Tell Us What You Did

A while back, ciphergoth posted an article on "picoeconomics", the theory that akrasia could be partially modeled as bargaining between present and future selves. I think the model is incomplete, because it doesn't explain how the analogy is instantiated in the real world, and I'd like to investigate that further sometime[1] - but it's a good first-order approximation.

For those of you too lazy to read the article (come on! It has pictures of naked people! Well, one naked person. Suspended from a graph of a hyperbolic curve) Ainslie argues that "intertemporal bargaining" is one way to overcome preference reversal. For example, an alcoholic has two conflicting preferences: right now, he would rather drink than not drink, but next year he would rather be the sort of person who never drinks than remain an alcoholic. But because his brain uses hyperbolic discounting, a process that pays more attention to his current utility than his future utility, he's going to hit the whiskey.

This sticks him in a sorites paradox. Honestly, it's not going to make much of a difference if he has one more drink, so why not hit the whiskey? Ainslie's answer is that he should set a hard-and-fast rule: "I will never drink alcohol". Following this rule will cure his alcoholism and help him achieve his dreams. He now has a very high preference for following the rule; a preference hopefully stronger than his current preference for whiskey.

Ainslie's other point is that this rule needs to really be hard-and-fast. If his rule is "I will drink less whiskey", then that leaves it open for him to say "Well, I'll drink some whiskey now, and none later; that counts as 'less'", and then the whole problem comes back just as bad as before. Likewise, if he says "It's my birthday, I'll let myself break the rule just this once," then soon he's likely to be saying "It's the Sunday before Cinco de Mayo, this calls for a celebration!" Ainslie has some much more formal and convincing ways of framing this, which is why you should read the article instead of just trusting this summary.

The stuff by Ainslie I read (I didn't spring for any of his dead-tree books) didn't offer any specific pointers for increasing your willpower[2], but it's pretty easy to read between the lines and figure out what applied picoeconomics ought to look like. In the interest of testing a scientific theory, not to mention the ongoing effort to take control of my own life, I've been testing picoeconomic techniques for the last two months.

The essence of picoeconomics is formally binding yourself to a rule with as few loopholes as possible. So the technique I decided to test[3] was to write out an oath detailing exactly what I wanted to do, list in nauseating detail all of the conditions under which I could or could not be released from this oath, and then bind myself to it, with the knowledge that if I succeeded I would have a great method of self-improvement and if I failed I would be dooming myself to a life of laziness forever (Ainslie's theories suggest that exaggeration is good in this case).

I chose a few areas of my life that I wanted to improve, of which the only one I want to mention in public is my poor study habits. I decided that I wanted to increase my current

study load from practically never looking at a book after school got out, up to two hours a day.

I wrote down - yes, literally wrote down - an oath in which I swore to study for two hours a day. I detailed exactly the conditions that would count as “studying” - no watching TV with an open book placed in my lap, for example.

I also included several release valves. The theory behind this was that if I simply broke the oath outright, the oath would no longer be credible and would lose its power (again, see Ainslie’s article), and there would be some point where I would be absolutely compelled to break the oath (for example, if a member of my family is in the emergency room, I refuse to read a book for an hour and a half before going to check up on them). I gave myself a whole bunch of cases in which I would be allowed to not study, guilt-free, and allowed myself five days a month when I could just take off studying for no reason (too tired, maybe). I also limited the original oath to a month, so that if it didn’t work I could adjust it without completely destroying the effectiveness of the oath forever. Finally, I swore the oath in a ceremonial fashion, calling upon various fictional deities for whom I have great respect.

One month later, I find that I kept to the terms of the oath exactly, which is no small achievement for me since my previous resolutions to study more have ended in apathy and failure. On an introspection level, the need to study each day felt exactly like the need to complete a project with a deadline, or to show up for work when the boss was expecting you. My brain clearly has different procedures for dealing with vague responsibilities it can weasel out of, and serious responsibilities it can’t, and the oath served to stick studying on the “serious” side of the line.

I am suitably cautious about other-optimizing and the typical mind fallacy, so I don't promise the same method will work for you. But I'd be interested to see if it did[4]. I'd be especially interested if everyone who tried it would post, right now, what they're trying so that in a month or so we can come back and see how many people kept their oath without having too much response bias.

Footnotes

1: I'm split on the value of picoeconomic theory. A lot of it seems either common-sense if taken as a vague model or metaphor, or obviously false if taken literally. But sometimes it's very good to have a formal model for common sense, and I'm optimistic about someone developing a more literal version of it that explains what's actually going on inside someone's head.

2: Ciphergoth, as far as you know does Ainslie ever start making practical suggestions based on his theory anywhere, or does he leave it entirely as an exercise for the reader?

3: I don't read a lot of stuff on productivity, so I might be reinventing the wheel here.

4: For people trying this, a few suggestions and caveats from my experience:

1. Do NOT make the oath open-ended. Set a time limit, and if you're happy at the end of that time limit, set another time limit.
2. Don't overdo it; this only works if you really do want the goal you're after more than you want momentary pleasure, people are notoriously bad at knowing what they want, and if you break an oath once you've set a

precedent and it'll be harder to keep a better-crafted oath next time. If I'd sworn six hours of studying a day, no way I'd have been able to keep it.

3. Set release valves.
4. Do something extremely measurable in which success or failure is a very yes-or-no affair, like how much time you do something for. Saying "study more" or "eat better" will be completely useless.
5. Read the article so you know the theory behind it and especially why it's important to always keep the rules.
6. Don't just think up the oath and figure it's in effect. Write it down and swear it aloud, more or less ceremonially, depending on your taste for drama and ritual.
7. Seriously, don't overdo it. Ego depletion and all that.

Schelling Fences on Slippery Slopes

Slippery slopes are themselves a slippery concept. Imagine trying to explain them to an alien:

"Well, we right-thinking people are quite sure that the Holocaust happened, so banning Holocaust denial would shut up some crackpots and improve the discourse. But it's one step on the road to things like banning unpopular political positions or religions, and we right-thinking people oppose that, so we won't ban Holocaust denial."

And the alien might well respond: "But you could just ban Holocaust denial, but not ban unpopular political positions or religions. Then you right-thinking people get the thing you want, but not the thing you don't want."

This post is about some of the replies you might give the alien.

Abandoning the Power of Choice

This is the boring one without any philosophical insight that gets mentioned only for completeness' sake. In this reply, giving up a certain point risks losing the ability to decide whether or not to give up other points.

For example, if people gave up the right to privacy and allowed the government to monitor all phone calls, online communications, and public places, then if someone launched a military coup, it would be very difficult to resist them because there would be no way to secretly organize a rebellion. This is also brought up in arguments about gun control a lot.

I'm not sure this is properly thought of as a slippery slope argument at all. It seems to be a more straightforward "Don't

give up useful tools for fighting tyranny" argument.

The Legend of Murder-Gandhi

Previously on Less Wrong's *The Adventures of Murder-Gandhi*: Gandhi is offered a pill that will turn him into an unstoppable murderer. He refuses to take it, because in his current incarnation as a pacifist, he doesn't want others to die, and he knows that would be a consequence of taking the pill. Even if we offered him $1 million to take the pill, his abhorrence of violence would lead him to refuse.

But suppose we offered Gandhi $1 million to take a different pill: one which would decrease his reluctance to murder by 1%. This sounds like a pretty good deal. Even a person with 1% less reluctance to murder than Gandhi is still pretty pacifist and not likely to go killing anybody. And he could donate the money to his favorite charity and perhaps save some lives. Gandhi accepts the offer.

Now we iterate the process: every time Gandhi takes the 1%-more-likely-to-murder-pill, we offer him another $1 million to take the same pill again.

Maybe original Gandhi, upon sober contemplation, would decide to accept $5 million to become 5% less reluctant to murder. Maybe 95% of his original pacifism is the only level at which he can be *absolutely sure* that he will still pursue his pacifist ideals.

Unfortunately, original Gandhi isn't the one making the choice of whether or not to take the 6th pill. 95%-Gandhi is. And 95% Gandhi doesn't care *quite* as much about pacifism as original Gandhi did. He still doesn't want to become a murderer, but it wouldn't be a disaster if he were just 90% as reluctant as original Gandhi, that stuck-up goody-goody.

What if there were a general principle that each Gandhi was comfortable with Gandhis 5% more murderous than himself, but no more? Original Gandhi would start taking the pills, hoping to get down to 95%, but 95%-Gandhi would start taking five more, hoping to get down to 90%, and so on until he's rampaging through the streets of Delhi, killing everything in sight.

Now we're tempted to say Gandhi shouldn't even take the first pill. But this also seems odd. Are we really saying Gandhi shouldn't take what's basically a free million dollars to turn himself into 99%-Gandhi, who might well be nearly indistinguishable in his actions from the original?

Maybe Gandhi's best option is to "fence off" an area of the slippery slope by establishing a Schelling point - an arbitrary point that takes on special value as a dividing line. If he can hold himself to the precommitment, he can maximize his winnings. For example, original Gandhi could swear a mighty oath to take only five pills - or if he didn't trust even his own legendary virtue, he could give all his most valuable possessions to a friend and tell the friend to destroy them if he took more than five pills. This would commit his future self to stick to the 95% boundary (even though that future self is itching to try to the same precommitment strategy to stick to its own 90% boundary).

Real slippery slopes will resemble this example if, each time we change the rules, we also end up changing our opinion about how the rules should be changed. For example, I think the Catholic Church may be working off a theory of "If we give up this traditional practice, people will lose respect for tradition and want to give up even more traditional practices, and so on."

Slippery Hyperbolic Discounting

One evening, I start playing *Sid Meier's Civilization* (IV, if you're wondering - V is terrible). I have work tomorrow, so I want to stop and go to sleep by midnight.

At midnight, I consider my alternatives. For the moment, I feel an urge to keep playing Civilization. But I know I'll be miserable tomorrow if I haven't gotten enough sleep. Being a hyperbolic discounter, I value the next ten minutes a lot, but after that the curve becomes pretty flat and maybe I don't value 12:20 much more than I value the next morning at work. Ten minutes' sleep here or there doesn't make any difference. So I say: "I will play Civilization for ten minutes - 'just one more turn' - and then I will go to bed."

Time passes. It is now 12:10. Still being a hyperbolic discounter, I value the next ten minutes a lot, and subsequent times much less. And so I say: I will play until 12:20, ten minutes sleep here or there not making much difference, and then sleep.

And so on until my empire bestrides the globe and the rising sun peeps through my windows.

This is pretty much the same process described above with Murder-Gandhi except that here the role of the value-changing pill is played by time and my own tendency to discount hyperbolically.

The solution is the same. If I consider the problem early in the evening, I can precommit to midnight as a nice round number that makes a good Schelling point. Then, when deciding whether or not to play after midnight, I can treat my decision not as "Midnight or 12:10" - because 12:10 will always win *that* particular race - but as "Midnight or abandoning the only

credible Schelling point and probably playing all night", which will be sufficient to scare me into turning off the computer.

(if I consider the problem at 12:01, I may be able to precommit to 12:10 if I am especially good at precommitments, but it's not a very natural Schelling point and it might be easier to say something like "as soon as I finish this turn" or "as soon as I discover this technology").

Coalitions of Resistance

Suppose you are a Zoroastrian, along with 1% of the population. In fact, along with Zoroastrianism your country has fifty other small religions, each with 1% of the population. 49% of your countrymen are atheist, and hate religion with a passion.

You hear that the government is considering banning the Taoists, who comprise 1% of the population. You've never liked the Taoists, vile doubters of the light of Ahura Mazda that they are, so you go along with this. When you hear the government wants to ban the Sikhs and Jains, you take the same tack.

But now you are in the unfortunate situation described by Martin Niemoller:

> *First they came for the socialists, and I did not speak out, because I was not a socialist.*
> *Then they came for the trade unionists, and I did not speak out, because I was not a trade unionist.*
> *Then they came for the Jews, and I did not speak out, because I was not a Jew.*
> *Then they came for me, but we had already abandoned the only defensible Schelling point*

With the banned Taoists, Sikhs, and Jains no longer invested in the outcome, the 49% atheist population has enough clout to ban Zoroastrianism and anyone else they want to ban. The better strategy would have been to have all fifty-one small religions form a coalition to defend one another's right to exist. In this toy model, they could have done so in an ecumenial congress, or some other literal strategy meeting.

But in the real world, there aren't fifty-one well-delineated religions. There are billions of people, each with their own set of opinions to defend. It would be impractical for everyone to physically coordinate, so they have to rely on Schelling points.

In the original example with the alien, I cheated by using the phrase "right-thinking people". In reality, figuring out who qualifies to join the Right-Thinking People Club is half the battle, and everyone's likely to have a different opinion on it. So far, the practical solution to the coordination problem, the "only defensible Schelling point", has been to just have everyone agree to defend everyone else without worrying whether they're right-thinking or not, and this is easier than trying to coordinate room for exceptions like Holocaust deniers. Give up on the Holocaust deniers, and no one else can be sure what other Schelling point you've committed to, if any…

…unless they can. In parts of Europe, they've banned Holocaust denial for years and everyone's been totally okay with it. There are also a host of other well-respected exceptions to free speech, like shouting "fire" in a crowded theater. Presumably, these exemptions are protected by tradition, so that they have become new Schelling points there, or are else so obvious that everyone except Holocaust deniers is willing to allow a special Holocaust denial exception without worrying it will impact their own case.

Summary

Slippery slopes legitimately exist wherever a policy not only affects the world directly, but affects people's willingness or ability to oppose future policies. Slippery slopes can sometimes be avoided by establishing a "Schelling fence" - a Schelling point that the various interest groups involved - or yourself across different values and times - make a credible precommitment to defend.

Democracy is the Worst Form of Government Except for All the Others Except Possibly Futarchy

I recently read Nate Silver's treatment of prediction markets in *The Signal and the Noise*. It was very good, but like most other such treatments it tended to focus on whether the best experts can do as well as prediction markets. The belief seems to be that if experts can equal - or perhaps even slightly outperform - these new market solutions, then no one can force us to switch to this complicated unorthodox system and we can safely keep relying on expert predictions. This post is about the reasons I disagree with that assessment.

A prediction market is a stock market analogue in which people buy and sell bets in order to predict the future. Someone creates a financial instrument that pays off $10 if the Democrats win the next election and $0 if they lose and people bid on the value of the instrument. If the instrument ends up priced at $6.50, that means the market thinks there's a 65% chance the Democrats will win.

These markets have some very interesting properties. The coolest is that it will always be the most consistently accurate source of information available. The proof is like so: suppose there were some source of information which was consistently better than the prediction market. In that case, whoever bet on the prediction market using that other source's predictions could consistently become rich. Many people like being rich, so someone would do this.

But this economic activity would move the prediction market's prices/predictions until they became as good as or better than

the other source's predictions. Unless you faster than everyone else playing the market, this will have already happened by the time you see its predictions. Therefore, the prediction market will always be the most consistently accurate source of information available.

Another cool property of prediction markets is that they're impossible to corrupt. Suppose the Democrats wanted to make themselves look more popular in order to convince campaign donors they were a shoo-in. So they spend a million dollars bidding on "The Democrats will win the next election" and driving the price up to $9.90, or a prediction of a 99% chance that the Democrats will win. It seems that the prediction market has been corrupted.

But suppose you notice this. You know the Democrats have a less than 99% chance of winning the election; therefore you can beat the prediction market. Therefore, you can get rich. You short shares of "The Democrats will win the election" until it goes down to whatever probability you think is correct. Other people do the same. Investors start a feeding frenzy as people realize what a big opportunity such an obviously wrong prediction is, and big firms with lots of money to spend on exactly this sort of situation join in. Eventually the prediction returns back to its correct level. The Democrats' plot to corrupt the market has turned into that the Democratic Party has turned into a plot to give away a million dollars by subsidizing more rational investors. The market easily returns to the correct level.

Robin Hanson has proposed that the government should use prediction markets to inform policy decisions. For example, one of the big controversies surrounding gun control is whether it will lower the crime rate (because fewer criminals have guns) or raise the crime rate (because fewer victims have

guns with which to defend themselves). In a futarchy, we would resolve this question by setting up a prediction market in which people predicted the future crime rate conditional upon gun control passing or failing. Since this would be the most accurate possible assessment of the evidence around guns and crime, we could use it to inform what legislation we wanted to pass.

So according to the conventional wisdom, this is a mildly interesting idea, but it depends a lot upon whether prediction markets can do better than the best of the experts who are already informing the debate on this subject. Nate Silver himself is a good example; he was, by many measures, more accurate than InTrade this election.

I am a big prediction market groupie, and I *don't care* whether top experts are a little better than prediction markets or vice versa. If you told me that Nate Silver can beat even a highly liquid prediction market by 5%, I would gain a little respect for Nate Silver but continue to push futarchy (government via prediction markets) over argentocracy (government by Nate Silver).

The reason is similar to the reason I (unlike a growing number of rationalists) continue to think democracy is a better system than monarchy, and it was most coherently explained by sci-fi writer/occasional antipope Charlie Stross.

The Mandate of Earth

We tend to think governments in general, and democracy in particular, should be optimized for good decision-making. To argue for democracy along these lines, one might suggest that democracy takes advantage of the wisdom of crowds, or that the population as a whole knows what it wants better than out of touch elites, or that monarchs would make bad decisions

because they're corrupt and interested only in their own power.

To argue *against* democracy along these lines, we might point out that elites are better-educated than the common people and less prone to populist arguments like "let's take resources from small unpopular groups and redistribute it to the majority". One might also just look around at how democratic judgments actually *work*. As @aristosophy puts it, "it was the 236th year of the reign of what would later be known as King The-American-People the Terrible."

Stross' argument for democracy says it was never intended as a means to optimize policy. It's got a few more modest goals, at which it succeeds admirably.

First, it's supposed to place an upper bound on how terrible a leader can be. America can do some stupid things sometimes, but we would never elect a Stalin, a Pol Pot, or a Kim Jong-Il - whereas military governments and monarchies often do end up with those kinds of people. One might counter-argue that a democracy elected Hitler, but this seems sufficiently explained by the majority of Hitler's badness focusing on minority groups without much voting power - a democracy would have trouble electing someone who was Hitler-level bad *towards the average voter*.

A democracy never has to worry about the crown prince being a psychotic bastard. It never has to worry about being forced to accept the last leader's feeble-minded son as the successor. I mean, we did it anyway, back in 2000. But we weren't *forced* to.

But second, and more important, a democracy provides a Schelling point. A Schelling point, recall, is an option which might or might not be the best, but which is not *too* bad and

which everyone agrees on in order to stop fighting. The President might not be the best leader. But he is very clearly *the* leader.

The importance of this cannot be overstated. The history of the world before democracy was a history of legitimacy squabbles. Some were succession squabbles - the king's psychotic younger son wants to seize the throne from the king's feeble-minded older son, or the Grand Vizier wants to murder the Sultan and start his own dynasty. Others were peasant revolts, where everyone just decides at the same time that they hate the king and decide to have a bloody civil war to overthrow him. Democracies get to avoid that.

In the six hundred fifty years between the Norman Conquest and the neutering of the English monarchy, Wikipedia lists about twenty revolts and civil wars, all the way from the Barons' Wars to the War of the Roses to the English Civil War. In the three hundred years since the neutering of the English monarchy and the switch to a more Parliamentary system, there have been exactly zero.

China is justly hailed as doing much better than the West with this because of their idea of the Mandate of Heaven, but even *they* collapsed into multiple feuding states around a dozen times in their history, for a total death toll in the tens of millions. *Romance of the Three Kingdoms*, which I reviewed recently, famously begins: "The empire, divided, seeks to unite; united, seeks to divide". For the vast majority of human history, there was this fatalism that there *was* going to be a civil war that destroyed your state, it was just a question of whether it happened tomorrow or next century.

In a non-democratic form of government, you're always going to have someone thinking they have more of a right to be in

charge than the guy who's there now. In a democracy, the criterion for legitimacy is an objective and easily verifiable one - they got the most votes in an election. If there's any dispute, you can just hold another election. As a Schelling point, it's hard to beat.

Yes, reactionaries, I totally just went there. I just said democracy was better than the Mandate of Heaven, *because it promotes stability*.

Prediction Markets Are Unimpeachable Experts

Democracy doesn't always perform optimally, but it always performs *fairly*. There are some biases in particular democracies, like the way the US primaries work, but the general concept of democracy is scrupulously fair, and that is enough to prevent people from starting civil wars.

Academia is different. Its state resembles that of pre-democratic governments, when anyone could choose a side, claim it was legitimate, and then get into endless protracted fights with the partisans of other sides. If you believe ObamaCare will destroy the economy, you will have no trouble finding a prestigious academic who agrees with you. Then all you need to do is accuse the other academics of bias, or cherry-picking, or using the wrong statistical test, or any of the other ways to discredit scientists you don't like (which are, to be fair, quite often true).

A democratic vote among the scientific establishment is insufficient to settle these topics. The most important problem is that it gives massive power to the people who determine who gets to be part of "the scientific establishment". A poll of theologians would establish that God exists; a poll of African Studies professors would establish that affirmative action is effective and morally obligatory; a poll of Sociology

professors would establish that capitalism is destroying the country and should be dismantled. Further, exactly which fields are biased in this way is itself a politically charged question: climate change deniers would argue that polling climatologists on global warming is exactly as messed up as polling theologians on God's existence.

This also creates overwhelming pressure for the government or special interest groups to take over scientific establishments. If we consider the intelligence community an "academic establishment" this is what happened during the Iraq War; the petrochemical industry is doing its best to subvert climatology and the pharmaceutical industry has quite a bit of power over medicine. If whether or not a drug worked was decided by a straight vote of all doctors, I bet the pharmaceutical companies would work a lot harder at gaining influence.

So not having any Schelling point - being hopelessly confused about the legitimacy of academic ideas - sucks. But a straight democratic vote of academics would also suck and be potentially unfair.

Prediction markets avoid these problems. There is no question of who the experts are: anyone can invest in a prediction market. There's no question of special interests taking it over; this just distributes free money to more honest investors.

Not only do they escape real bias, but more importantly they escape *perceived* bias. It is breathtakingly beautiful how impossible it is to rail that a prediction market is the tool of the liberal media or whatever. You just tell Limbaugh: "Wait, you think the prediction market has a consistently liberal bias? Then invest on the conservative sides of issues and you get free money for having discovered this startling economic

fact!" If Limbaugh invests his fortune and turns out to be right, he's laughing all the way to the bank *and* improving the system. If Limbaugh *claims* the market is biased but refuses to invest it in, everyone knows he's just spouting hot air.

Nate Silver might do better than a prediction market, I don't know. But Nate Silver is not a Schelling point. Nobody chose him as Official Statistics Guy via a fair process. And if someone objected to his beliefs, they could accuse him of bias and he would have no recourse until it was too late.

If a prediction market is *almost* as good as Nate, and it is also unbiased and impossible to accuse of bias, we have our Schelling point. Barack Obama can say something like "Obamacare won't be unaffordable, in fact it will *cut* the size of the budget deficit!" And if Rush Limbaugh says "You're lying, or relying on data collected by liberal hacks", Obama can just retort "No, seriously, the prediction market says there's an 80% chance I'm right", and Limbaugh will just have to admit he's right and slink away.

Just as democracy made it harder to fight over leadership, prediction markets make it *harder to fight over beliefs*. We can still fight over values, of course - if you hate teenagers having sex, and I don't care about it, we can debate that all day long. But if we want to know whether a certain law will raise the pregnancy rate, there will be only one correct answer, and it will only be a mouse-click away.

I think this would have more positive effects than anyone anticipates. If people took it seriously, not only would the gun control debate be over in an hour, but it would end on the objectively right side, *whichever side that was*. If single-payer would be better than Obamacare, we could implement single-payer and anyone who tried to make up horror stories about

how it would destroy health care would be laughed out of the room. And once these issues have gone away, maybe we can reach the point where half the country stops hating the other half because of disagreements which are largely over factual issues.

Right now we're going backward from this future. A prediction market has to be very liquid (ie have many users spending lots of time on it) before it becomes any good at predicting things. But the US government just cracked down on the largest prediction market, InTrade, because they classify it as "online gambling". This has much reduced its liquidity and set the entire field back by years. IARPA, a government intelligence agency thing, has a toy prediction market going, but it's much more limited without real money.

I hope that someone soon starts a bitcoin prediction market outside the US government's reach. It might fail - prediction market users and bitcoin users are both small minorities, and the disjunction of two small minorities might be too small to provide the necessary liquidity - but maybe later when dollar-bitcoin convertibility becomes more fluid, its time will come. This is the sort of idea I would totally pursue myself if I had money, time, technical knowledge, business acumen, entrepreneurial spirit, legal advice, and about twenty other abstract and concrete resources I do not possess. As it is I sort of fantasize about making enough money in medicine to fund someone who has the other nineteen.

Eight Short Studies on Excuses

The Clumsy Game-Player

You and a partner are playing an Iterated Prisoner's Dilemma. Both of you have publicly pre-committed to the tit-for-tat strategy. By iteration 5, you're going happily along, raking up the bonuses of cooperation, when your partner unexpectedly presses the "defect" button.

"Uh, sorry," says your partner. "My finger slipped."

"I still have to punish you just in case," you say. "I'm going to defect next turn, and we'll see how you like it."

"Well," said your partner, "knowing that, I guess I'll defect next turn too, and we'll both lose out. But hey, it was just a slipped finger. By not trusting me, you're costing us both the benefits of one turn of cooperation."

"True", you respond "but if I don't do it, you'll feel free to defect whenever you feel like it, using the 'finger slipped' excuse."

"How about this?" proposes your partner. "I promise to take extra care that my finger won't slip again. You promise that if my finger does slip again, you will punish me terribly, defecting for a bunch of turns. That way, we trust each other again, and we can still get the benefits of cooperation next turn."

You don't believe that your partner's finger really slipped, not for an instant. But the plan still seems like a good one. You accept the deal, and you continue cooperating until the experimenter ends the game.

After the game, you wonder what went wrong, and whether you could have played better. You decide that there was no better way to deal with your partner's "finger-slip" - after all, the plan you enacted gave you maximum possible utility under the circumstances. But you wish that you'd pre-committed, at the beginning, to saying "and I will punish finger slips equally to deliberate defections, so make sure you're careful."

The Lazy Student

You are a perfectly utilitarian school teacher, who attaches exactly the same weight to others' welfare as to your own. You have to have the reports of all fifty students in your class ready by the time midterm grades go out on January 1st. You don't want to have to work during Christmas vacation, so you set a deadline that all reports must be in by December 15th or you won't grade them and the students will fail the class. Oh, and your class is Economics 101, and as part of a class project all your students have to behave as selfish utility-maximizing agents for the year.

It costs your students 0 utility to turn in the report on time, but they gain +1 utility by turning it in late (they enjoy procrastinating). It costs you 0 utility to grade a report turned in before December 15th, but -30 utility to grade one after December 15th. And students get 0 utility from having their reports graded on time, but get -100 utility from having a report marked incomplete and failing the class.

If you say "There's no penalty for turning in your report after deadline," then the students will procrastinate and turn in their reports late, for a total of +50 utility (1 per student times fifty students). You will have to grade all fifty reports during Christmas break, for a total of - 1500 utility (-30 per report times fifty reports). Total utility is -1450.

So instead you say “If you don’t turn in your report on time, I won’t grade it.” All students calculate the cost of being late, which is +1 utility from procrastinating and -100 from failing the class, and turn in their reports on time. You get all reports graded before Christmas, no students fail the class, and total utility loss is zero. Yay!

Or else - one student comes to you the day after deadline and says “Sorry, I was really tired yesterday, so I really didn’t want to come all the way here to hand in my report. I expect you’ll grade my report anyway, because I know you to be a perfect utilitarian, and you’d rather take the -30 utility hit to yourself than take the -100 utility hit to me.”

You respond “Sorry, but if I let you get away with this, all the other students will turn in their reports late in the summer.” She says “Tell you what - our school has procedures for changing a student’s previously given grade. If I ever do this again, or if I ever tell anyone else about this, you can change my grade to a fail. Now you know that passing me this one time won’t affect anything in the future. It certainly can’t affect the past. So you have no reason not to do it.” You believe her when she says she’ll never tell, but you say “You made this argument because you believed me to be the sort of person who would accept it. In order to prevent other people from making the same argument, I have to be the sort of person who wouldn’t accept it. To that end, I’m going to not accept your argument.”

The Grieving Student

A second student comes to you and says “Sorry I didn’t turn in my report yesterday. My mother died the other day, and I wanted to go to her funeral.”

You say “Like all economics professors, I have no soul, and so am unable to sympathize with your loss. Unless you can make an argument that would apply to all rational actors in my position, I can’t grant you an extension.”

She says “If you did grant this extension, it wouldn’t encourage other students to turn in their reports late. The other students would just say ‘She got an extension because her mother died’. They know they won’t get extensions unless they kill their own mothers, and even economics students aren’t that evil. Further, if you don’t grant the extension, it won’t help you get more reports in on time. Any student would rather attend her mother’s funeral than pass a course, so you won’t be successfully motivating anyone else to turn in their reports early.”

You think for a while, decide she’s right, and grant her an extension on her report.

The Sports Fan

A third student comes to you and says “Sorry I didn’t turn in my report yesterday. The Bears’ big game was on, and as I’ve told you before, I’m a huge Bears fan. But don’t worry! It’s very rare that there’s a game on this important, and not many students here are sports fans anyway. You’ll probably never see a student with this exact excuse again. So in a way, it’s not that different from the student here just before me, the one whose mother died.”

You respond “It may be true that very few people will be able to say both that they’re huge Bears fans, and that there’s a big Bears game on the day before the report comes due. But by accepting your excuse, I establish a precedent of accepting excuses that are *approximately this good*. And there are many other excuses approximately as good as yours. Maybe

someone's a big soap opera fan, and the season finale is on the night before the deadline. Maybe someone loves rock music, and there's a big rock concert on. Maybe someone's brother is in town that week. Practically anyone can come up with an excuse as good as yours, so if I accept your late report, I have to accept everyone's.

"The student who was here before you, that's different. We, as a society, already have an ordering in which a family member's funeral is one of the most important things around. By accepting her excuse, I'm establishing a precedent of accepting any excuse approximately that good, but almost no one will ever have an excuse that good. Maybe a few people who are really sick, someone struggling with a divorce or a breakup, that kind of thing. Not the hordes of people who will be coming to me if I give you your exemption."

The Murderous Husband

You are the husband of a wonderful and beautiful lady whom you love very much - and whom you just found in bed with another man. In a rage, you take your hardcover copy of Introduction To Game Theory and knock him over the head with it, killing him instantly (it's a pretty big book).

At the murder trial, you plead to the judge to let you go free. "Society needs to lock up murderers, as a general rule. After all, they are dangerous people who cannot be allowed to walk free. However, I only killed that man because he was having an affair with my wife. In my place, anyone would have done the same. So the crime has no bearing on how likely I am to murder someone else. I'm not a risk to anyone who isn't having an affair with my wife, and after this incident I plan to divorce and live the rest of my days a bachelor. Therefore, you

have no need to deter me from future murders, and can safely let me go free."

The judge responds: "You make a convincing argument, and I believe that you will never kill anyone else in the future. However, other people will one day be in the position you were in, where they walk in on their wives having an affair. Society needs to have a credible pre-commitment to punishing them if they succumb to their rage, in order to deter them from murder."

"No," you say, "I understand your reasoning, but it won't work. If you've never walked in on your wife having an affair, you can't possibly understand the rage. No matter how bad the deterrent was, you'd still kill the guy."

"Hm," says the judge. "I'm afraid I just can't believe anyone could ever be quite that irrational. But I see where you're coming from. I'll give you a lighter sentence."

The Bellicose Dictator

You are the dictator of East Examplestan, a banana republic subsisting off its main import, high quality hypothetical scenarios. You've always had it in for your ancestral enemy, West Examplestan, but the UN has made it clear that any country in your region that aggressively invades a neighbor will be severely punished with sanctions and possible enforced "regime change." So you decide to leave the West alone for the time being.

One day, a few West Examplestanis unintentionally wander over your unmarked border while prospecting for new scenario mines. You immediately declare it a "hostile incursion" by "West Examplestani spies", declare war, and take the Western capital in a sneak attack.

The next day, Ban Ki-moon is on the phone, and he sounds angry. “I thought we at the UN had made it perfectly clear that countries can’t just invade each other anymore!”

“But didn’t you read our propaganda mouthpi…ahem, official newspaper? We didn’t *just* invade. We were responding to Western aggression!”

“Balderdash!” says the Secretary-General. “Those were a couple of lost prospectors, and you know it!”

“Well,” you say. “Let’s consider your options. The UN needs to make a credible pre-commitment to punish aggressive countries, or everyone will invade their weaker neighbors. And you’ve got to follow through on your threats, or else the pre-commitment won’t be credible anymore. But you don’t actually like following through on your threats. Invading rogue states will kill a lot of people on both sides and be politically unpopular, and sanctions will hurt your economy *and* lead to heart-rending images of children starving. What you’d really like to do is let us off, but in a way that doesn’t make other countries think they’ll get off too.

“Luckily, we can make a credible story that we were following international law. Sure, it may have been stupid of us to mistake a few prospectors for an invasion, but there’s no international law against being stupid. If you dismiss us as simply misled, you don’t have to go through the trouble of punishing us, and other countries won’t think they can get away with anything.

“Nor do you need to live in fear of us doing something like this again. We’ve already demonstrated that we won’t go to war without a casus belli. If other countries can refrain from giving us one, they have nothing to fear.”

Ban Ki-moon doesn't believe your story, but the countries that would bear the economic brunt of the sanctions and regime change decide they believe it just enough to stay uninvolved.

The Peyote-Popping Native

You are the governor of a state with a large Native American population. You have banned all mind-altering drugs, with the honorable exceptions of alcohol, tobacco, caffeine, and several others, because you are a red-blooded American who believes that they would drive teenagers to commit crimes.

A representative of the state Native population comes to you and says: "Our people have used peyote religiously for hundreds of years. During this time, we haven't become addicted or committed any crimes. Please grant us a religious exemption under the First Amendment to continue practicing our ancient rituals." You agree.

A leader of your state's atheist community breaks into your office via the ventilation systems (because seriously, how else is an atheist leader going to get access to a state governor?) and says: "As an atheist, I am offended that you grant exemptions to your anti-peyote law for religious reasons, but not for, say, recreational reasons. This is unfair discrimination in favor of religion. The same is true of laws that say Sikhs can wear turbans in school to show support for God, but my son can't wear a baseball cap in school to show support for the Yankees. Or laws that say Muslims can get time off state jobs to pray five times a day, but I can't get time off my state job for a cigarette break. Or laws that say state functions will include special kosher meals for Jews, but not special pasta meals for people who really like pasta."

You respond "Although my policies may seem to be saying religion is more important than other potential reasons for

breaking a rule, one can make a non-religious case justifying them. One important feature of major world religions is that their rituals have been fixed for hundreds of years. Allowing people to break laws for religious reasons makes religious people very happy, but does not weaken the laws. After all, we all know the few areas in which the laws of the major US religions as they are currently practiced conflict with secular law, and none of them are big deals. So the general principle 'I will allow people to break laws if it is necessary to established and well-known religious rituals' is relatively low-risk and makes people happy without threatening the concept of law in general. But the general principle 'I will allow people to break laws for recreational reasons' is *very* high risk, because it's sufficient justification for almost anyone breaking any law."

"I would love to be able to serve everyone the exact meal they most wanted at state dinners. But if I took your request for pasta because you liked pasta, I would have to follow the general principle of giving everyone the meal they most like, which would be prohibitively expensive. By giving Jews kosher meals, I can satisfy a certain particularly strong preference without being forced to satisfy anyone else's."

The Well-Disguised Atheist

The next day, the atheist leader comes in again. This time, he is wearing a false mustache and sombrero. "I represent the Church of Driving 50 In A 30 Mile Per Hour Zone," he says. "For our members, going at least twenty miles per hour over the speed limit is considered a sacrament. Please grant us a religious exemption to traffic laws."

You decide to play along. "How long has your religion existed, and how many people do you have?" you ask.

"Not very long, and not very many people," he responds.

“I see,” you say. “In that case, you’re a cult, and not a religion at all. Sorry, we don’t deal with cults.”

“What, exactly, is the difference between a cult and a religion?”

“The difference is that cults have been formed recently enough, and are small enough, that we are suspicious of them existing for the purpose of taking advantage of the special place we give religion. Granting an exemption for your cult would challenge the credibility of our pre-commitment to punish people who break the law, because it would mean anyone who wants to break a law could just found a cult dedicated to it.”

“How can my cult become a real religion that deserves legal benefits?”

“You’d have to become old enough and respectable enough that it becomes implausible that it was created for the purpose of taking advantage of the law.”

“That sounds like a lot of work.”

“Alternatively, you could try writing awful science fiction novels and hiring a ton of lawyers. I hear that also works these days.”

Conclusion

In all these stories, the first party wants to credibly pre-commit to a rule, but also has incentives to forgive other people’s deviations from the rule. The second party breaks the rules, but comes up with an excuse for why its infraction should be forgiven.

The first party’s response is based not only on whether the person’s excuse is believable, not even on whether the person’s excuse is morally valid, but on whether the excuse

can be accepted without straining the credibility of their previous pre-commitment.

The general principle is that by accepting an excuse, a rule-maker is also committing themselves to accepting all equally good excuses in the future. There are some exceptions - accepting an excuse in private but making sure no one else ever knows, accepting an excuse once with the express condition that you will never accept any other excuses - but to some degree these are devil's bargains, as anyone who can predict you will do this can take advantage of you.

These stories give an idea of excuses different from the one our society likes to think it uses, namely that it accepts only excuses that are true and that reflect well upon the character of the person giving the excuse. I'm not saying that the common idea of excuses doesn't have value - but I think the game theory view also has some truth to it. I also think the game theoretic view can be useful in cases where the common view fails. It can inform cases in law, international diplomacy, and politics where a tool somewhat stronger than the easily-muddled common view is helpful.

Revenge as Charitable Act

Someone on Reddit told a story about his job as a convenience store cashier. One day a known problem customer walked in, bought an item with a $10 bill, then said he'd paid with a $50 and demanded $40 extra change. The cashier was on to the ploy and politely refused. The customer called in the manager, who proceeded to chew out the cashier for arguing with customers and ordered him to hand over the man's $40. At the end of the day, surprise surprise, the cashier ended up $40 short. The manager got angry and, over the cashier's protests, docked him $40 in pay - most of his earnings for the day - because of "his" mistake.

I have a short temper at times, and when I read this my blood boiled. If trying to reason with the guy didn't work, I could totally see myself yelling at my manager right there and then in front of the whole store, telling him how unfair and incompetent he was, quitting on the spot in the hopes that it screwed up the store's business for the new few weeks, taking the guy to small claims court, and seeing if I could get a newspaper to take up the story and drag this guy's name through the mud as much as possible.

Most people and ideologies who claim wisdom, including most of the world's religions, condemn that sort of thing as "seeking revenge". They point out, not unreasonably, that this would hurt both myself and my manager. My manager would be out an employee and have a court case and an angry mob of newspaper-readers to deal with. As for me, I'd be out of a job, forced to do a lot of work getting the court case and newspaper article together, and probably known around town as the guy who threw a fit over an employment squabble - and none of it

would get me my $40 back. So the conventional wisdom is to turn the other cheek, forgive the manager, and keep everyone happy.

Economists would take an opposite view: they would say that my revenge is a self-sacrificing act of charity. Imagine a world in which everyone who was swindled by a crappy employer quit immediately and went on jihad against them. The world would very soon be empty of crappy employers; the only successful employers would be those who realized they couldn't get away with mistreating their workers. By taking revenge, I'm sacrificing my own pleasure - my job and my time - in order to help create a world where crappy behavior isn't tolerated and doesn't happen anymore.

(lest this sound like I'm arguing for communism or something, the same applies to other common forms of revenge: road rage for bad driving, spitting in customers' food for being rude to waitstaff.)

Now someone's going to come in and say that the *most* moral thing to do is keep the job so I can donate the money I get from it to charity, and okay, point well taken. But assuming I'm not going to do that, it seems the *more* moral thing to do is to take as much revenge as possible. And religion and spirituality are usually really on board about this "self-sacrifice for the sake of the community" thing, which makes it odd for them to so vehemently be against it.

I don't really like this conclusion because the beautiful "hold no ill towards anyone, just be serene and forgiving" ethic of the religions appeals more to my own aesthetic. But the logic seems hard to escape.

(it does seem to be a big deal that people feel slighted more often than they actually have been, and so a world where

people always took revenge would involve a lot of extraneous revenge-seeking for imagined offenses. But that's hardly the least convenient possible world, and there are times when I can be pretty certain I've been genuinely slighted).

Would Your Real Preferences Please Stand Up?

Related to: Cynicism in Ev Psych and Econ

In Finding the Source, a commenter says:

> I have begun wondering whether claiming to be victim of 'akrasia' might just be a way of admitting that your real preferences, as revealed in your actions, don't match the preferences you want to signal (believing what you want to signal, even if untrue, makes the signals more effective).

I think I've seen Robin put forth ~~something like this argument~~ [EDIT: Something related, but very different], and TGGP points out that Brian Caplan explicitly believes pretty much the same thing[1]:

> I've previously argued that much - perhaps most - talk about "self-control" problems reflects social desirability bias rather than genuine inner conflict.
>
> Part of the reason why people who spend a lot of time and money on socially disapproved behaviors say they "want to change" is that that's what they're supposed to say.
>
> Think of it this way: A guy loses his wife and kids because he's a drunk. Suppose he sincerely prefers alcohol to his wife and kids. He still probably won't admit it, because people judge a sinner even more harshly if he is unrepentent. The drunk who says "I was such a fool!" gets some pity; the drunk who says "I like Jack

> Daniels better than my wife and kids" gets horrified looks. And either way, he can keep drinking.

I'll call this the Cynic's Theory of Akrasia, as opposed to the Naive Theory. I used to think it was plausible. Now that I think about it a little more, I find it meaningless. Here's what changed my mind.

What part of the mind, exactly, prefers a socially unacceptable activity (like drinking whiskey or browsing Reddit) to an acceptable activity (like having a wife and kids, or studying)? The conscious mind? As Bill said in his comment, it doesn't seem like it works this way. I've had akrasia myself, and I never consciously think "Wow, I really like browsing Reddit… but I'll trick everyone else into thinking I'd rather be studying so I get more respect. Ha ha! The fools will never see it coming!"

No, my conscious mind fully believes that I would rather be studying[2]. And this even gets reflected in my actions. I've tried anti-procrastination techniques, both successfully and unsuccessfully, without ever telling them to another living soul. People trying to diet don't take out the cupcakes as soon as no one else is looking (or, if they do, they feel guilty about it).

This is as it should be. It is a classic finding in evolutionary psychology: the person who wants to fool others begins by fooling themselves. Some people even call the conscious mind the "public relations officer" of the brain, and argue that its entire point is to sit around and get fooled by everything we want to signal. As Bill said, "believing the signals, even if untrue, makes the signals more effective."

Now we have enough information to see why the Cynic's Theory is equivalent to the Naive Theory.

The Naive Theory says that you really want to stop drinking, but some force from your unconscious mind is hijacking your actions. The Cynic's Theory says that you really want to keep drinking, but your conscious mind is hijacking your thoughts and making you think otherwise.

In both cases, the conscious mind determines the signal and the unconscious mind determines the action. The only difference is which preference we define as "real" and worthy of sympathy. In the Naive Theory, we sympathize with the conscious mind, and the *problem* is the unconscious mind keeps committing contradictory actions. In the Cynic's Theory, we symapthize with the unconscious mind, and the *problem* is the conscious mind keeps sending out contradictory signals. The Naive say: find some way to make the unconscious mind stop hijacking actions! The Cynic says: find some way to make the conscious mind stop sending false signals!

So why prefer one theory over the other? Well, I'm not surprised that it's mostly economists who support the Cynic's Theory. Economists are understandably interested in revealed preferences[3], because revealed preferences are revealed by economic transactions and are the ones that determine the economy. It's perfectly reasonable for an economist to care only about those and dimiss any other kind of preference as a red herring that has to be removed before economic calculations can be done. Someone like a philosopher, who is more interested in thought and the mind, might be more susceptible to the identify-with-conscious-thought Naive Theory.

But notice how the theory you choose also has serious political implications[4]. Consider how each of the two ways of looking at the problem would treat this example:

> A wealthy liberal is a member of many environmental organizations, and wants taxes to go up to pay for better conservation programs. However, she can't bring herself to give up her gas-guzzling SUV, and is usually too lazy to sort all her trash for recycling.

I myself throw my support squarely behind the Naive Theory. Conscious minds are potentially rational[5], informed by morality, and qualia-laden. Unconscious minds aren't, so who cares what they think?

Footnotes:

1: Caplan says that the lack of interest in Stickk offers support for the Cynic's Theory, but I don't see why it should, unless we believe the mental balance of power should be different when deciding whether to use Stickk than when deciding whether to do anything else.

Caplan also suggests in another article that he has never experienced procrastination as akrasia. Although I find this surprising, I don't find it absolutely impossible to believe. His mind may either be exceptionally well-integrated, or it may send signals differently. It seems within the range of normal human mental variation.

2: Of course, I could be lying here, to signal to you that I have socially acceptable beliefs. I suppose I can only make my point if you often have the same experience, or if you've caught someone else fighting akrasia when they didn't know you were there.

3: Even the term "revealed preferences" imports this value system, as if the act of buying something is a revelation that

drives away the mist of the false consciously believed preferences.

4: For a real-world example of a politically-charged conflict surrounding the question of whether we should judge on conscious or unconscious beliefs, see Robin's post Redistribution Isn't About Sympathy and my reply.

5: Differences between the conscious and unconscious mind should usually correspond to differences between the goals of a person and the "goals" of the genome, or else between subgoals important today and subgoals important in the EEA.

Are Wireheads Happy?

Related to: Utilons vs. Hedons, Would Your Real Preferences Please Stand Up

And I don't mean that question in the semantic "but what is happiness?" sense, or in the deep philosophical "but can anyone not facing struggle and adversity truly be happy?" sense. I mean it in the totally literal sense. Are wireheads having fun?

They look like they are. People and animals connected to wireheading devices get upset when the wireheading is taken away and will do anything to get it back. And it's electricity shot directly into the reward center of the brain. What's not to like?

Only now neuroscientists are starting to recognize a difference between "reward" and "pleasure", or call it "wanting" and "liking". The two are usually closely correlated. You want something, you get it, then you feel happy. The simple principle behind our entire consumer culture. But do neuroscience and our own experience really support that?

It would be too easy to point out times when people want things, get them, and then later realize they weren't so great. That could be a simple case of misunderstanding the object's true utility. What about wanting something, getting it, realizing it's not so great, and then wanting it just as much the next day? Or what about not wanting something, getting it, realizing it makes you very happy, and then continuing not to want it?

The first category, "things you do even though you don't like them very much" sounds like many drug addictions. Smokers may enjoy smoking, and they may want to avoid the

physiological signs of withdrawl, but neither of those is enough to explain their reluctance to quit smoking. I don't smoke, but I made the mistake of starting a can of Pringles yesterday. If you asked me my favorite food, there are dozens of things I would say before "Pringles". Right now, and for the vast majority of my life, I feel no desire to go and get Pringles. But once I've had that first chip, my motivation for a second chip goes through the roof, without my subjective assessment of how tasty Pringles are changing one bit.

Think of the second category as "things you procrastinate even though you like them." I used to think procrastination applied only to things you disliked but did anyway. Then I tried to write a novel. I loved writing. Every second I was writing, I was thinking "This is so much fun". And I never got past the second chapter, because I just couldn't motivate myself to sit down and start writing. Other things in this category for me: going on long walks, doing yoga, reading fiction. I can know with near certainty that I will be happier doing X than Y, and still go and do Y.

Neuroscience provides some basis for this. A University of Michigan study analyzed the brains of rats eating a favorite food. They found separate circuits for "wanting" and "liking", and were able to knock out either circuit without affecting the other (it was actually kind of cute - they measured the number of times the rats licked their lips as a proxy for "liking", though of course they had a highly technical rationale behind it). When they knocked out the "liking" system, the rats would eat exactly as much of the food without making any of the satisifed lip-licking expression, and areas of the brain thought to be correlated with pleasure wouldn't show up in the MRI. Knock out "wanting", and the rats seem to enjoy the food as

much when they get it but not be especially motivated to seek it out. To quote the science[1]:

> Pleasure and desire circuitry have intimately connected but distinguishable neural substrates. Some investigators believe that the role of the mesolimbic dopamine system is not primarily to encode pleasure, but "wanting" i.e. incentive-motivation. On this analysis, endomorphins and enkephalins - which activate mu and delta opioid receptors most especially in the ventral pallidum - are most directly implicated in pleasure itself. Mesolimbic dopamine, signalling to the ventral pallidum, mediates desire. Thus "dopamine overdrive", whether natural or drug-induced, promotes a sense of urgency and a motivation to engage with the world, whereas direct activation of mu opioid receptors in the ventral pallidum induces emotionally self-sufficient bliss.

The wanting system is activated by dopamine, and the liking system is activated by opioids. There are enough connections between them that there's a big correlation in their activity, but the correlation isn't one and in fact activation of the opioids is less common than the dopamine. Another quote:

> It's relatively hard for a brain to generate pleasure, because it needs to activate different opioid sites together to make you like something more. It's easier to activate desire, because a brain has several 'wanting' pathways available for the task. Sometimes a brain will like the rewards it wants. But other times it just wants them.

So you could go through all that trouble to find a black market brain surgeon who'll wirehead you, and you'll end up not even

being happy. You'll just really really want to keep the wirehead circuit running.

Problem: large chunks of philosophy and economics are based upon wanting and liking being the same thing.

By definition, if you choose X over Y, then X is a higher utility option than Y. That means utility represents wanting and not liking. ~~But good utilitarians (and, presumably, artificial intelligences) try to maximize utility~~ (or do they?). This correlates contingently with maximizing happiness, but not necessarily. In a worst-case scenario, it might not correlate at all - two possible such scenarios being wireheading and an AI without the appropriate common sense.

Thus the deep and heavy ramifications. A more down-to-earth example came to mind when I was reading something by Steven Landsburg recently (not recommended). I don't have the exact quote, but it was something along the lines of:

> According to a recent poll, two out of three New Yorkers say that, given the choice, they would rather live somewhere else. But all of them have the choice, and none of them live anywhere else. A proper summary of the results of this poll would be: two out of three New Yorkers lie on polls.

This summarizes a common strain of thought in economics, the idea of "revealed preferences". People tend to say they like a lot of things, like family or the environment or a friendly workplace. Many of the same people who say these things then go and ignore their families, pollute, and take high-paying but stressful jobs. The traditional economic explanation is that the people's actions reveal their true preferences, and that all the talk about caring about family and the environment is just

stuff people say to look good and gain status. If a person works hard to get lots of money, spends it on an iPhone, and doesn't have time for their family, the economist will say that this proves that they value iPhones more than their family, no matter what they may say to the contrary.

The difference between enjoyment and motivation provides an argument that could rescue these people. It may be that a person really does enjoy spending time with their family more than they enjoy their iPhone, but they're more motivated to work and buy iPhones than they are to spend time with their family. If this were true, people's introspective beliefs and public statements about their values would be true as far as it goes, and their tendency to work overtime for an iPhone would be as much a "hijacking" of their "true preferences" as a revelation of them. This accords better with my introspective experience, with happiness research, and with common sense than the alternative.

Not that the two explanations are necessarily entirely contradictory. One could come up with a story about how people are motivated to act selfishly but enjoy acting morally, which allows them to tell others a story about how virtuous they are while still pursuing their own selfish gain.

Go too far toward the liking direction, and you risk something different from wireheading only in that the probe is stuck in a different part of the brain. Go too far in the wanting direction, and you risk people getting lots of shiny stuff they thought they wanted but don't actually enjoy. So which form of good should altruists, governments, FAIs, and other agencies in the helping people business respect?

Sources/Further Reading:

1. Wireheading.com, especially on a particular University of Michigan study

2. New York Times: A Molecule of Motivation, Dopamine Excels at its Task

3. Slate: The Powerful and Mysterious Brain Circuitry…

4. Related journal articles (1, 2, 3)

Guilt: Another Gift Nobody Wants

Evolutionary psychology has made impressive progress in understanding the origins of morality. Along with the many posts about these origins on Less Wrong I recommend Robert Wright's *The Moral Animal* for an excellent introduction to the subject.

Guilt does not naturally fall out of these explanations. One can imagine a mind design that although often behaving morally for the same reasons we do, sometimes decides a selfish approach is best and pursues that approach without compunction. In fact, this design would have advantages; it would remove a potentially crippling psychological burden, prevent loss of status from admission of wrongdoing, and allow more rational calculation of when moral actions are or are not advantageous. So why guilt?

In one of the few existing writings I could find on the subject, Tooby and Cosmides theorize that "guilt functions as an emotion mode specialized for recalibration of regulatory variables that control trade-offs in welfare between self and other."

If I understand their meaning, they are saying that when an action results in a bad outcome, guilt is a byproduct of updating your mental processes so that it doesn't happen again. In their example, if you don't share food with your sister, and your sister starves and becomes sick, your brain gives you a strong burst of negative emotion around the event so that you reconsider your decision not to share. It is generally a bad idea to disagree with Tooby and Cosmides, but this explanation doesn't satisfy me for several reasons.

First, guilt is just as associated with good outcomes as bad outcomes. If I kill my brother so I can inherit the throne, then even if everything goes according to plan and I become king, I may still feel guilt. But why should I recalibrate here? My original assumptions - that fratricide would be easy and useful - were entirely correct. But I am still likely to feel bad about it. In fact, some criminals report feeling "relieved" when caught, as if a negative outcome decreased their feelings of guilt instead of exacerbating them.

Second, guilt is not only an emotion, but an entire complex of behaviors. Our modern word self-flagellation comes from the old practice of literally whipping one's self out of feelings of guilt or unworthiness. We may not literally self-flagellate anymore, but when I feel guilty I am less likely to do activities I enjoy and more likely to deliberately make myself miserable.

Third, although guilt can be very private it has an undeniable social aspect. People have messaged me at 3 AM in the morning just to tell me how guilty they feel about something they did to someone I've never met; this sort of outpouring of emotion can even be therapeutic. The aforementioned self-flagellators would parade around town in their sackcloth and ashes, just in case anyone didn't know how guilty they felt. And we expect guilt in certain situations: a criminal who feels guilty about what ey has done may get a shorter sentence.

Fourth, guilt sometimes occurs even when a person has done nothing wrong. People who through no fault of their own are associated with disasters can nevertheless report "survivor's guilt" and feel like events were partly their fault. If this is a tool for recalibrating choices, it is a very bad one. This is not a knockdown argument - a lot of mental adaptations are very bad at what they do - but it should at least raise suspicion that there is another part to the puzzle besides recalibration.

THE PARABLE OF THE LAWYER

Suppose you need a lawyer for some important and very lucrative legal case. And suppose by a freak legislative oversight, your state has no laws against legal malpractice and unethical lawyers can get off scot-free. You are going to want to invest a lot of effort into evaluating the morals of the many lawyers anxious to take your case.

One lawyer you meet, Mr. Dewey, has an unusual appearance. A small angel, about the size of a rat, sits on his right shoulder holding an electric cattle prod. This is remarkable, and so you remark upon it.

Mr. Dewey scowls. "That angel has been sitting there for as long as I can remember," he tells you. "Every time I do something wrong, she pokes me with her prod. If it's a minor sin like profanity, maybe she'll only poke me once or twice, but if I lie or swindle, she'll turn the power up on max and keep shocking me for days. It's a miserable, miserable existence, and I'm constantly scared to death I'll slip up and make her angry, but I can't figure out how to get rid of her."

You express some skepticism about this story, so Mr. Dewey offers to demonstrate. He says a mild curse word, and sure enough, the angel pokes him with the cattle prod, giving him a mild electric shock.

Suddenly, Mr. Dewey is a very attractive candidate for your lucrative case. You can be assured that he won't swindle you, because whatever gains he might take from the swindle are less attractive than the punishment he would get from the angel afterwards.

Surgeon Paul Brand considered pain so useful to the body's functioning that he called it "the gift nobody wants". Mr.

Dewey's angel is also such a gift, even though he might not appreciate it: clients worried about ethical issues will bring their patronage to his law firm, giving him a major advantage over the competition.

Whereas normally we must trust a lawyer's altruism if we expect em not to con us, in Mr. Dewey's case we need only trust him to pursue his own self-interest. This, then, is the role of guilt: it provides assurance to others that we will be punished for our misdeeds even if there is no external authority to punish us, avoiding Parfitian hitchhiker dilemmas and ensuring fair play. The assurance of punishment ensures fair play and makes mutually beneficial transactions possible.

FAKEABLE AND UNFAKEABLE SIGNALS

The big difference between Mr. Dewey and ourselves is that where Mr. Dewey has unquestionable evidence of his commitment to self punishment in the form of a very visible angel on his shoulder, for the rest of us guilt is a private mental affair and can be faked. It would seem to be a winning strategy, then, to claim a tendency to guilt while not really having one.

Ms. Wolfram is Mr. Dewey's main competitor, and is outraged at her rival's business success. In an attempt to even the scales, she buys a plastic angel figure from the local church and glues it to her shoulder. "Look!" she tells clients. "I, too, suffer pain when I commit misdeeds!" Her business shoots up to the same high levels as Mr. Dewey's.

One day, the news comes that Mr. Dewey was spotted whipping himself in the town square. When asked why, he explained that in a moment of weakness, he had overcharged a customer. His angel, who had lost its cattle prod, was mind-

controlling him into the self-flagellation in place of its more usual punishment.

This provides an impressive bar for Ms. Wolfram to live up to. Sure, she could just whip herself like Mr. Dewey is doing. But it wouldn't be worth it - she just doesn't like the money enough that she would whip herself after every swindle just to drum up business. If she's going to have to whip herself to fake remorse whenever she commits wrongdoing, her best policy really *is* to genuinely stop swindling people.

Mr. Dewey has found an unfakeable signal. Even though whipping himself in public is one of the most unpleasant things he could do, in this case it is good business practice. It once again differentiates him from Ms. Wolfram and restores his status as the city's most desirable attorney.

In evolutionary terms, guilt becomes more credible the more it requires publicly visible behavior that no reasonable cheat would want to fake. Hurting oneself, avoiding pleasurable activities, lowering your own status, and withdrawing from social activities are all evolutionary costly and therefore good ways to prove you are experiencing guilt; the usual vocal, postural, and facial cues of being miserable are also useful.

There's no reason people should evolve an all-consuming sense of guilt. If an opportunity comes along where the benefits of cheating are greater than the social costs, an organism should still take it. Therefore, guilt has to be unpleasant but not infinitely unpleasant. A person who committed suicide in response to even the slightest moral infraction would be trustworthy, but they'd miss out if an excellent opportunity to win major gains for cheating happened to fall into their lap.

The conspicuous experience of guilt is an evolutionarily advantageous way of assuring potential trading partners that you will be punished for defection. The behaviors associated with guilt are costly signals that help differentiate false claims of guilt from the real thing and add to public verifiability of the punishment involved.

UNDESERVED GUILT

If you kill your brother in order to inherit the throne, you probably deserve whatever guilt you feel. But in the phenomenon of "survivor's guilt", people feel guilt for events that weren't even remotely their fault. Maybe you go hiking with your brother, and through no fault of your own he trips and falls down a crevasse and dies, and now you feel guilty. Why?

Hunter-gatherer societies were more violent than our own; statistics differ but by some estimates around 30% of hunter-gatherer males died of homicide. Even as late as the Bronze Age, Biblical figures who killed their brothers comprise a rather impressive list including Cain, Solomon, Ammon, Abimelech, and Jehoram; Jacob's sons merely attempted to do so. So the priors for suspicious death must have been very different in the olden days.

Further, in such a crime-ridden culture, there may have been more incentives to blame an enemy for a death, even if that enemy was not responsible. A person whose brother has accidentally died on a hiking trip with no witnesses would be very targetable.

And even in less drastic situations than blaming survivors for a death, there may be other possible threats to reputation. If there is only one survivor of a battle, he may be suspected of

cowardice; if there is only one survivor of a disaster, she may be suspected of running away without helping others.

Therefore, it would be advantageous to have a method of proving your innocence. Suppose that you would gain benefits X from killing your brother and covering it up, but that you would suffer losses Y if you were suspected of the crime and punished. A precommitment to a policy of experiencing a level of guilt between X and Y provides a tool for proving your innocence. It would no longer be in your self-interest to kill your brother, because you will suffer so much guilt that you won't be able to enjoy the benefits of your crime; your would-be accusers realize this and admit your innocence, saving you from the still worse outcome Y.

In this case, guilt would be an entirely adaptive response to a disaster with which you were associated, even if your own actions were beyond reproach. A level of unhappiness worse than any benefits you could get by profiting the tragedy, but less than any punishment you might receive if you were suspected of profiting from the tragedy, would be helpful in clearing your name of any wrongdoing.

(The proposed mechanism is almost identical to one cited in Thornhill and Palmer's controversial and unpleasant evolutionary account of post-traumatic stress after rape.)

This theory makes some testable predictions, which as far as I know have not been tested:

- People should feel guiltier about events for which reasonable suspicion might exist that they played a part; for example, if your brother slipped and fell while you were hiking alone with him rather than in a large group with many witnesses.
- People should feel guiltier about events for which they might profit; for example, if you stood to inherit money from your

brother, or never liked him much anyway.
- People may be suspicious of people who come out of a disaster feeling no survivor's guilt.

CONCLUSION

Guilt, like pain, is "a gift nobody wants". Because people with guilt are known to punish themselves for moral wrongdoing, their social group considers them more trustworthy and they gain the advantages of trade and cooperation. In order to prove that their guilt is real rather than feigned, they use costly signals like deliberate self-harm and self-denial to display their punishment publicly

When one has done nothing wrong, it can sometimes be advantageous to paradoxically display guilt in order to prove one's lack of wrongdoing. These costly signals demonstrate that it is not in one's self-interest to lie about these matters, while still being less costly than the punishment for defection.

Although this could theoretically be mediated by the behavioral strategies of a sufficiently intelligent and Machiavellian unconscious mind, it fits within the framework of evolutionary psychology and can also be interpreted in evolutionary terms.

VII. Cognition and Association

Diseased Thinking: Dissolving Questions about Disease

Related to: Disguised Queries, Words as Hidden Inferences, Dissolving the Question, Eight Short Studies on Excuses

> *Today's therapeutic ethos, which celebrates curing and disparages judging, expresses the liberal disposition to assume that crime and other problematic behaviors reflect social or biological causation. While this absolves the individual of responsibility, it also strips the individual of personhood, and moral dignity*
>
> — *George Will, townhall.com*

Sandy is a morbidly obese woman looking for advice.

Her husband has no sympathy for her, and tells her she obviously needs to stop eating like a pig, and would it kill her to go to the gym once in a while?

Her doctor tells her that obesity is primarily genetic, and recommends the diet pill orlistat and a consultation with a surgeon about gastric bypass.

Her sister tells her that obesity is a perfectly valid lifestyle choice, and that fat-ism, equivalent to racism, is society's way of keeping her down.

When she tells each of her friends about the opinions of the others, things really start to heat up.

Her husband accuses her doctor and sister of absolving her of personal responsibility with feel-good platitudes that in the end will only prevent her from getting the willpower she needs to start a real diet.

Her doctor accuses her husband of ignorance of the real causes of obesity and of the most effective treatments, and accuses her sister of legitimizing a dangerous health risk that could end with Sandy in hospital or even dead.

Her sister accuses her husband of being a jerk, and her doctor of trying to medicalize her behavior in order to turn it into a "condition" that will keep her on pills for life and make lots of money for Big Pharma.

Sandy is fictional, but similar conversations happen every day, not only about obesity but about a host of other marginal conditions that some consider character flaws, others diseases, and still others normal variation in the human condition. Attention deficit disorder, internet addiction, social anxiety disorder (as one skeptic said, didn't we used to call this "shyness"?), alcoholism, chronic fatigue, oppositional defiant disorder ("didn't we used to call this being a teenager?"), compulsive gambling, homosexuality, Aspergers' syndrome, antisocial personality, even depression have all been placed in two or more of these categories by different people.

Sandy's sister may have a point, but this post will concentrate on the debate between her husband and her doctor, with the understanding that the same techniques will apply to evaluating her sister's opinion. The disagreement between Sandy's husband and doctor centers around the idea of "disease". If obesity, depression, alcoholism, and the like are diseases, most people default to the doctor's point of view; if they are not diseases, they tend to agree with the husband.

The debate over such marginal conditions is in many ways a debate over whether or not they are "real" diseases. The usual surface level arguments trotted out in favor of or against the proposition are generally inconclusive, but this post will apply a

host of techniques previously discussed on Less Wrong to illuminate the issue.

What is Disease?

In Disguised Queries , Eliezer demonstrates how a word refers to a cluster of objects related upon multiple axes. For example, in a company that sorts red smooth translucent cubes full of vanadium from blue furry opaque eggs full of palladium, you might invent the word “rube” to designate the red cubes, and another “blegg”, to designate the blue eggs. Both words are useful because they “carve reality at the joints” - they refer to two completely separate classes of things which it’s practically useful to keep in separate categories. Calling something a “blegg” is a quick and easy way to describe its color, shape, opacity, texture, and chemical composition. It may be that the odd blegg might be purple rather than blue, but in general the characteristics of a blegg remain sufficiently correlated that “blegg” is a useful word. If they weren’t so correlated - if blue objects were equally likely to be palladium-containing-cubes as vanadium-containing-eggs, then the word “blegg” would be a waste of breath; the characteristics of the object would remain just as mysterious to your partner after you said “blegg” as they were before.

“Disease”, like “blegg”, suggests that certain characteristics always come together. A rough sketch of some of the characteristics we expect in a disease might include:

1. Something caused by the sorts of thing you study in biology: proteins, bacteria, ions, viruses, genes.
2. Something involuntary and completely immune to the operations of free will
3. Something rare; the vast majority of people don’t have it
4. Something unpleasant; when you have it, you want to get rid of it

5. Something discrete; a graph would show two widely separate populations, one with the disease and one without, and not a normal distribution.
6. Something commonly treated with science-y interventions like chemicals and radiation.

Cancer satisfies every one of these criteria, and so we have no qualms whatsoever about classifying it as a disease. It's a type specimen, the sparrow as opposed to the ostrich. The same is true of heart attack, the flu, diabetes, and many more.

Some conditions satisfy a few of the criteria, but not others. Dwarfism seems to fail (5), and it might get its status as a disease only after studies show that the supposed dwarf falls way out of normal human height variation. Despite the best efforts of transhumanists, it's hard to convince people that aging is a disease, partly because it fails (3). Calling homosexuality a disease is a poor choice for many reasons, but one of them is certainly (4): it's not necessarily unpleasant.

The marginal conditions mentioned above are also in this category. Obesity arguably sort-of-satisfies criteria (1), (4), and (6), but it would be pretty hard to make a case for (2), (3), and (5).

So, is obesity really a disease? Well, is Pluto really a planet? Once we state that obesity satisfies some of the criteria but not others, it is meaningless to talk about an additional fact of whether it "really deserves to be a disease" or not.

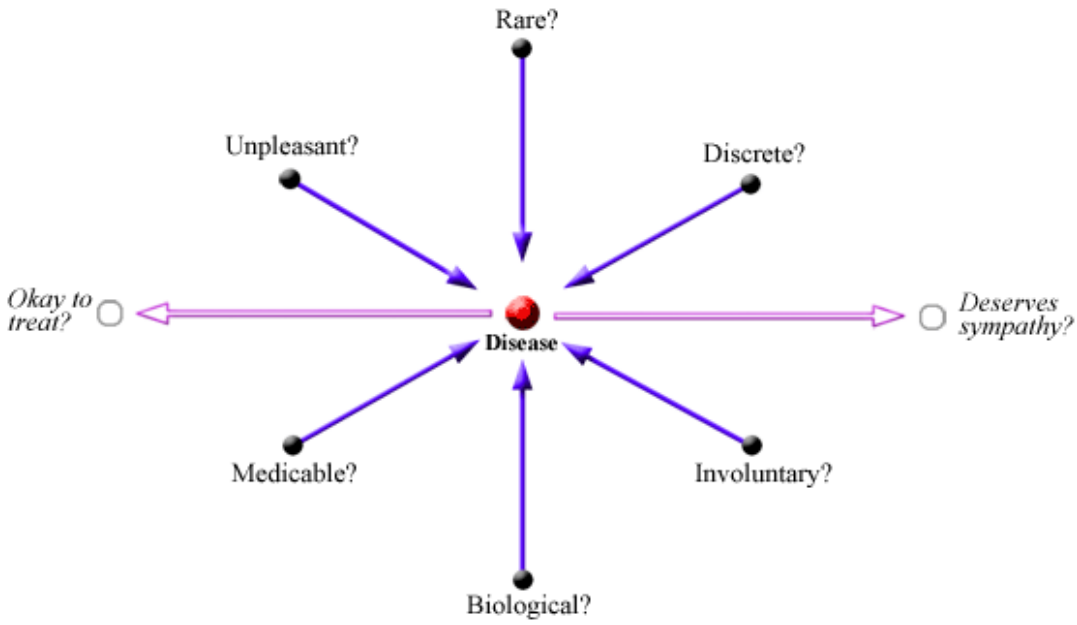
If it weren't for those pesky hidden inferences…

Hidden Inferences From Disease Concept

The state of the disease node, meaningless in itself, is used to predict several other nodes with non-empirical content. In English: we make value decisions based on whether we call something a "disease" or not.

If something is a real disease, the patient deserves our sympathy and support; for example, cancer sufferers must universally be described as "brave". If it is not a real disease, people are more likely to get our condemnation; for example Sandy's husband who calls her a "pig" for her inability to control her eating habits. The difference between "shyness" and "social anxiety disorder" is that people with the first get called "weird" and told to man up, and people with the second get special privileges and the sympathy of those around them.

And if something is a real disease, it is socially acceptable (maybe even mandated) to seek medical treatment for it. If it's not a disease, medical treatment gets derided as a "quick fix" or an "abdication of personal responsibility". I have talked to several doctors who are uncomfortable suggesting gastric bypass surgery, even in people for whom it is medically indicated, because they believe it is morally wrong to turn to medicine to solve a character issue.



While a condition's status as a "real disease" ought to be meaningless as a "hanging node" after the status of all other nodes have been determined, it has acquired political and

philosophical implications because of its role in determining whether patients receive sympathy and whether they are permitted to seek medical treatment.

If we can determine whether a person should get sympathy, and whether they should be allowed to seek medical treatment, independently of the central node "disease" or of the criteria that feed into it, we will have successfully unasked the question "are these marginal conditions real diseases" and cleared up the confusion.

Sympathy or Condemnation?

Our attitudes toward people with marginal conditions mainly reflect a deontologist libertarian (libertarian as in "free will", not as in "against government") model of blame. In this concept, people make decisions using their free will, a spiritual entity operating free from biology or circumstance. People who make good decisions are intrinsically good people and deserve good treatment; people who make bad decisions are intrinsically bad people and deserve bad treatment. But people who make bad decisions for reasons that are outside of their free will may not be intrinsically bad people, and may therefore be absolved from deserving bad treatment. For example, if a normally peaceful person has a brain tumor that affects areas involved in fear and aggression, they go on a crazy killing spree, and then they have their brain tumor removed and become a peaceful person again, many people would be willing to accept that the killing spree does not reflect negatively on them or open them up to deserving bad treatment, since it had biological and not spiritual causes.

Under this model, deciding whether a condition is biological or spiritual becomes very important, and the rationale for worrying over whether something "is a real disease" or not is plain to see. Without figuring out this extremely difficult question, we are at risk of either blaming people for things they don't deserve, or

else letting them off the hook when they commit a sin, both of which, to libertarian deontologists, would be terrible things. But determining whether marginal conditions like depression have a spiritual or biological cause is difficult, and no one knows how to do it reliably.

Determinist consequentialists can do better. We believe it's biology all the way down. Separating spiritual from biological illnesses is impossible and unnecessary. Every condition, from brain tumors to poor taste in music, is "biological" insofar as it is encoded in things like cells and proteins and follows laws based on their structure.

But determinists don't just ignore the very important differences between brain tumors and poor taste in music. Some biological phenomena, like poor taste in music, are encoded in such a way that they are extremely vulnerable to what we can call social influences: praise, condemnation, introspection, and the like. Other biological phenomena, like brain tumors, are completely immune to such influences. This allows us to develop a more useful model of blame.

The consequentialist model of blame is very different from the deontological model. Because all actions are biologically determined, none are more or less metaphysically blameworthy than others, and none can mark anyone with the metaphysical status of "bad person" and make them "deserve" bad treatment. Consequentialists don't on a primary level want anyone to be treated badly, full stop; thus is it written: "Saddam Hussein doesn't deserve so much as a stubbed toe." But if consequentialists don't believe in punishment for its own sake, they do believe in punishment for the sake of, well, consequences. Hurting bank robbers may not be a good in and of itself, but it will prevent banks from being robbed in the future. And, one might infer, although alcoholics may not deserve

condemnation, societal condemnation of alcoholics makes alcoholism a less attractive option.

So here, at last, is a rule for which diseases we offer sympathy, and which we offer condemnation: if giving condemnation instead of sympathy decreases the incidence of the disease enough to be worth the hurt feelings, condemn; otherwise, sympathize. Though the rule is based on philosophy that the majority of the human race would disavow, it leads to intuitively correct consequences. Yelling at a cancer patient, shouting "How dare you allow your cells to divide in an uncontrolled manner like this; is that the way your mother raised you??!" will probably make the patient feel pretty awful, but it's not going to cure the cancer. Telling a lazy person "Get up and do some work, you worthless bum," very well might cure the laziness. The cancer is a biological condition immune to social influences; the laziness is a biological condition susceptible to social influences, so we try to socially influence the laziness and not the cancer.

The question "Do the obese deserve our sympathy or our condemnation," then, is asking whether condemnation is such a useful treatment for obesity that its utility outweights the disutility of hurting obese people's feelings. This question may have different answers depending on the particular obese person involved, the particular person doing the condemning, and the availability of other methods for treating the obesity, which brings us to…

The Ethics of Treating Marginal Conditions

If a condition is susceptible to social intervention, but an effective biological therapy for it also exists, is it okay for people to use the biological therapy instead of figuring out a social solution? My gut answer is "Of course, why wouldn't it be?", but apparently lots of people find this controversial for some reason.

In a libertarian deontological system, throwing biological solutions at spiritual problems might be disrespectful or dehumanizing, or a band-aid that doesn't affect the deeper problem. To someone who believes it's biology all the way down, this is much less of a concern.

Others complain that the existence of an easy medical solution prevents people from learning personal responsibility. But here we see the status-quo bias at work, and so can apply a preference reversal test. If people really believe learning personal responsibility is more important than being not addicted to heroin, we would expect these people to support deliberately addicting schoolchildren to heroin so they can develop personal responsibility by coming off of it. Anyone who disagrees with this somewhat shocking proposal must believe, on some level, that having people who are not addicted to heroin is more important than having people develop whatever measure of personal responsibility comes from kicking their heroin habit the old-fashioned way.

But the most convincing explanation I have read for why so many people are opposed to medical solutions for social conditions is a signaling explanation by Robin Hans…wait! no!…by Katja Grace. On her blog, she says:

> *…the situation reminds me of a pattern in similar cases I have noticed before. It goes like this. Some people make personal sacrifices, supposedly toward solving problems that don't threaten them personally. They sort recycling, buy free range eggs, buy fair trade, campaign for wealth redistribution etc. Their actions are seen as virtuous. They see those who don't join them as uncaring and immoral. A more efficient solution to the problem is suggested. It does not require personal sacrifice. People who have not previously sacrificed support it. Those who have previously*

> *sacrificed object on grounds that it is an excuse for people to get out of making the sacrifice. The supposed instrumental action, as the visible sign of caring, has become virtuous in its own right. Solving the problem effectively is an attack on the moral people.*

A case in which some people eat less enjoyable foods and exercise hard to avoid becoming obese, and then campaign against a pill that makes avoiding obesity easy demonstrates some of the same principles.

There are several very reasonable objections to treating any condition with drugs, whether it be a classical disease like cancer or a marginal condition like alcoholism. The drugs can have side effects. They can be expensive. They can build dependence. They may later be found to be placebos whose efficacy was overhyped by dishonest pharmaceutical advertising.. They may raise ethical issues with children, the mentally incapacitated, and other people who cannot decide for themselves whether or not to take them. But these issues do not magically become more dangerous in conditions typically regarded as "character flaws" rather than "diseases", and the same good-enough solutions that work for cancer or heart disease will work for alcoholism and other such conditions (but see here).

I see no reason why people who want effective treatment for a condition should be denied it or stigmatized for seeking it, whether it is traditionally considered "medical" or not.

Summary

People commonly debate whether social and mental conditions are real diseases. This masquerades as a medical question, but its implications are mainly social and ethical. We use the concept of disease to decide who gets sympathy, who gets blame, and who gets treatment.

Instead of continuing the fruitless “disease” argument, we should address these questions directly. Taking a determinist consequentialist position allows us to do so more effectively. We should blame and stigmatize people for conditions where blame and stigma are the most useful methods for curing or preventing the condition, and we should allow patients to seek treatment whenever it is available and effective.

The Noncentral Fallacy — The Worst Argument in the World?

Related to: Leaky Generalizations, Replace the Symbol With The Substance, Sneaking In Connotations

David Stove once ran a contest to find the Worst Argument In The World, but he awarded the prize to his own entry, and one that shored up his politics to boot. It hardly seems like an objective process.

If he can unilaterally declare a Worst Argument, then so can I. I declare the Worst Argument In The World to be this: "X is in a category whose archetypal member gives us a certain emotional reaction. Therefore, we should apply that emotional reaction to X, even though it is not a central category member."

Call it the Noncentral Fallacy. It sounds dumb when you put it like that. Who even does that, anyway?

It sounds dumb only because we are talking soberly of categories and features. As soon as the argument gets framed in terms of *words*, it becomes so powerful that somewhere between many and most of the bad arguments in politics, philosophy and culture take some form of the noncentral fallacy. Before we get to those, let's look at a simpler example.

Suppose someone wants to build a statue honoring Martin Luther King Jr. for his nonviolent resistance to racism. An opponent of the statue objects: "But Martin Luther King was a *criminal*!"

Any historian can confirm this is correct. A criminal is technically someone who breaks the law, and King knowingly

broke a law against peaceful anti-segregation protest - hence his famous Letter from Birmingham Jail.

But in this case calling Martin Luther King a criminal is the noncentral. The archetypal criminal is a mugger or bank robber. He is driven only by greed, preys on the innocent, and weakens the fabric of society. Since we don't like these things, calling someone a "criminal" naturally lowers our opinion of them.

The opponent is saying "Because you don't like criminals, and Martin Luther King is a criminal, you should stop liking Martin Luther King." But King doesn't share the important criminal features of being driven by greed, preying on the innocent, or weakening the fabric of society that made us dislike criminals in the first place. Therefore, even though he is a criminal, there is no reason to dislike King.

This all seems so nice and logical when it's presented in this format. Unfortunately, it's also one hundred percent contrary to instinct: the urge is to respond "Martin Luther King? A criminal? No he wasn't! You take that back!" This is why the noncentral is so successful. As soon as you do that you've fallen into their trap. Your argument is no longer about whether you should build a statue, it's about whether King was a criminal. Since he was, you have now lost the argument.

Ideally, you should just be able to say "Well, King was the good kind of criminal." But that seems pretty tough as a debating maneuver, and it may be even harder in some of the cases where the noncentral Fallacy is commonly used.

Now I want to list some of these cases. Many will be political[1], for which I apologize, but it's hard to separate out a bad argument from its specific instantiations. None of these

examples are meant to imply that the position they support is wrong (and in fact I myself hold some of them). They only show that certain particular arguments for the position are flawed, such as:

"Abortion is *murder*!" The archetypal murder is Charles Manson breaking into your house and shooting you. This sort of murder is bad for a number of reasons: you prefer not to die, you have various thoughts and hopes and dreams that would be snuffed out, your family and friends would be heartbroken, and the rest of society has to live in fear until Manson gets caught. If you define murder as "killing another human being", then abortion is technically murder. But it has none of the downsides of murder Charles Manson style. Although you can criticize abortion for many reasons, insofar as "abortion is murder" is an invitation to apply one's feelings in the Manson case directly to the abortion case, it ignores the latter's lack of the features that generated those intuitions in the first place[2].

"**Genetic engineering to cure diseases is *eugenics*!**" Okay, you've got me there: since eugenics means "trying to improve the gene pool" that's clearly right. But what's wrong with eugenics? "What's wrong with eugenics? Hitler did eugenics! Those unethical scientists in the 1950s who sterilized black women without their consent did eugenics!" "And what was wrong with what Hitler and those unethical scientists did?" "What do you mean, what was wrong with them? Hitler killed millions of people! Those unethical scientists ruined people's lives." "And does using genetic engineering to cure diseases kill millions of people, or ruin anyone's life?" "Well…not really." "Then what's wrong with it?" "It's *eugenics*!"

"Evolutionary psychology is *sexist*!" If you define "sexist" as "believing in some kind of difference between the sexes",

this is true of at least some evo psych. For example, Bateman's Principle states that in species where females invest more energy in producing offspring, mating behavior will involve males pursuing females; this posits a natural psychological difference between the sexes. "Right, so you admit it's sexist!" "And why exactly is sexism bad?" "Because sexism claims that men are better than women and that women should have fewer rights!" "Does Bateman's principle claim that men are better than women, or that women should have fewer rights?" "Well…not really." "Then what's wrong with it?" "It's *sexist!"*

A second, subtler use of the noncentral fallacy goes like this: "X is in a category whose archetypal member gives us an emotional reaction. Therefore, we should apply that same emotional reaction to X even if X gives some benefit that outweighs the harm."

"Capital punishment is *murder*!" Charles Manson-style murder is solely harmful. This kind of murder produces really strong negative feelings. The proponents of capital punishment believe that it might decrease crime, or have some other attending benefits. In other words, they believe it's "the good kind of murder"[3], just like the introductory example concluded that Martin Luther King was "the good kind of criminal". But since normal murder is so taboo, it's really hard to take the phrase "the good kind of murder" seriously, and just mentioning the word "murder" can call up exactly the same amount of negative feelings we get from the textbook example.

"Affirmative action is *racist*!" True if you define racism as "favoring certain people based on their race", but once again, our immediate negative reaction to the archetypal example of racism (the Ku Klux Klan) cannot be generalized to an immediate negative reaction to affirmative action. Before we

generalize it, we have to check first that the problems that make us hate the Ku Klux Klan (violence, humiliation, divisiveness, lack of a meritocratic society) are still there. Then, even if we do find that some of the problems persist (like disruption of meritocracy, for example) we have to prove that it doesn't produce benefits that outweigh these harms.

"**Taxation is *theft*!**" True if you define theft as "taking someone else's money regardless of their consent", but though the archetypal case of theft (breaking into someone's house and stealing their jewels) has nothing to recommend it, taxation (arguably) does. In the archetypal case, theft is both unjust and socially detrimental. Taxation keeps the first disadvantage, but arguably subverts the second disadvantage if you believe being able to fund a government has greater social value than leaving money in the hands of those who earned it. The question then hinges on the relative importance of these disadvantages. Therefore, you can't dismiss taxation without a second thought just because you have a natural disgust reaction to theft in general. You would also have to prove that the supposed benefits of this form of theft don't outweigh the costs.

Now, because most arguments are rapid-fire debate-club style, sometimes it's still useful to say "Taxation isn't theft!" At least it beats saying "Taxation is theft but nevertheless good", then having the other side say "Apparently my worthy opponent thinks that theft can be good; we here on this side would like to bravely take a stance *against* theft", and then having the moderator call time before you can explain yourself. If you're in a debate club, do what you have to do. But if you have the luxury of philosophical clarity, you would do better to forswear the Dark Arts and look a little deeper into what's going on.

Are there ever cases in which this argument pattern can be useful? Yes. For example, it may be a groping attempt to suggest a Schelling fence; for example, a principle that one must never commit theft even when it would be beneficial because that would make it harder to distinguish and oppose the really bad kinds of theft. Or it can be an attempt to spark conversation by pointing out a potential contradiction: for example "Have you noticed that taxation really does contain some of the features you dislike about more typical instances of theft? Maybe you never even thought about that before? Why do your moral intuitions differ in these two cases? Aren't you being kind of hypocritical?" But this usage seems pretty limited - once your interlocutor says "Yes, I considered that, but the two situations are different for reasons X, Y, and Z" the conversation needs to move on; there's not much point in continuing to insist "But it's *theft*!"

But in most cases, I think this is more of an *emotional* argument, or even an argument from "You would look silly saying that". You really *can't* say "Oh, he's the good kind of criminal", and so if you have a potentially judgmental audience and not much time to explain yourself, you're pretty trapped. You have been forced to round to the archetypal example of that word and subtract exactly the information that's most relevant.

But in all other cases, the proper response to being asked to subtract relevant information is "No, why should I?" - and that's why this is the worst argument in the world.

Footnotes

1: On advice from the community, I have deliberately included three mostly-liberal examples and three-mostly conservative

examples, so save yourself the trouble of counting them up and trying to speculate on this article's biases.

2: This should be distinguished from deontology, the belief that there is some provable moral principle about how you can never murder. I don't think this is *too* important a point to make, because only a tiny fraction of the people who debate these issues have thought that far ahead, and also because my personal and admittedly controversial opinion is that much of deontology is just an attempt to formalize and justify this fallacy.

3: Some people "solve" this problem by saying that "murder" only refers to "non-lawful killing", which is exactly as creative a solution as redefining "criminal" to mean "person who breaks the law and is not Martin Luther King." Identifying the noncentral fallacy is a more complete solution: for example, it covers the related (mostly sarcastic) objection that "imprisonment is kidnapping".

4: EDIT 8/2013: I've edited this article a bit after getting some feedback and complaints. In particular I tried to remove some LW jargon which turned off some people who were being linked to this article but were unfamiliar with the rest of the site.

5: EDIT 8/2013: The other complaint I kept getting is that this is an uninteresting restatement of some other fallacy (no one can agree which, but poisoning the well comes up particularly often). The question doesn't seem too interesting to me - I never claimed particular originality, a lot of fallacies blend into each other, and the which-fallacy-is-which game isn't too exciting anyway - but for the record I don't think it is. Poisoning the well is a presentation of two different facts, such as "Martin Luther King was a plagiarist…oh, by the way, what

do you think of Martin Luther King's civil rights policies?" It may have no relationship to categories, and it's usually something someone else does to you as a conscious rhetorical trick. Noncentral fallacy is presenting a single fact, but using category information to frame it in a misleading way - and it's often something people do to themselves. The above plagiarism example of poisoning the well is *not* noncentral fallacy. If you think this essay is about bog-standard poisoning the well, then either there is an alternative meaning to poisoning the well I'm not familiar with, or you are missing the point.

The Power of Positivist Thinking

Related to: No Logical Positivist I, Making Beliefs Pay Rent, How An Algorithm Feels From Inside, Disguised Queries

Call me non-conformist, call me one man against the world, but…I kinda like logical positivism.

The logical positivists were a dour, no-nonsense group of early 20th-century European philosophers. Indeed, the phrase "no-nonsense" seems almost invented to describe the Positivists. They liked nothing better then to reject the pet topics of other philosophers as being untestable and therefore meaningless. Is the true also the beautiful? Meaningless! Is there a destiny to the affairs of humankind? Meaningless? What is justice? Meaningless! Are rights inalienable? Meaningless!

Positivism became stricter and stricter, defining more and more things as meaningless, until someone finally pointed out that positivism itself was meaningless by the positivists' definitions, at which point the entire system vanished in a puff of logic. Okay, it wasn't that simple. It took several decades and Popper's falsifiabilism to seal its coffin. But vanish it did. It remains one of the least lamented theories in the history of philosophy, because if there is one thing philosophers hate it's people telling them they can't argue about meaningless stuff.

But if we've learned anything from fantasy books, it is that any cabal of ancient wise men destroyed by their own hubris at the height of their glory must leave behind a single ridiculously powerful artifact, which in the right hands gains the power to dispel darkness and annihilate the forces of evil.

The positivists left us the idea of verifiability, and it's time we started using it more.

Eliezer, in No Logical Positivist I, condemns the positivist notion of verifiability for excluding some perfectly meaningful propositions. For example, he says, it may be that a chocolate cake formed in the center of the sun on 8/1/2008, then disappeared after one second. This statement seems to be meaningful; that is, there seems to be a difference between it being true or false. But there's no way to test it (at least without time machines and sundiver ships, which we can't prove are possible) so the logical positivists would dismiss it as nonsense.

I am not an expert in logical positivism; I have two weeks studying positivism in an undergrad philosophy class under my belt, and little more. If Eliezer says that is how the positivists interpreted their verifiability criterion, I believe him. But it's not the way I would have done things, if I'd been in 1930s Vienna. I would have said that any statement corresponding to a state of the material universe, reducible in theory to things like quarks and photons, testable by a being who has access to the machine running the universe[1] and who can check the logs at will - such a statement is meaningful[2]. In this case the chocolate cake example passes: it corresponds to a state of the material world, and is clearly visible on the universe's logs. "Rights are inalienable" remains meaningless, however. At the risk of reinventing the wheel[3], I will call this interpretation "soft positivism".

My positivism gets even softer, though. Consider the statement "Google is a successful company." Though my knowledge of positivism is shaky, I believe that most positivists would reject this as meaningless; "success" is too fuzzy to be reduced to anything objective. But if positivism is true, it should add up to normality: we shouldn't find that an obviously useful statement like "Google is a successful company" is total

nonsense. I interpret the statement to mean certain objectively true propositions like "The average yearly growth rate for Google has been greater than the average yearly growth rate for the average company", which itself reduces down to a question of how much money Google made each year, which is something that can be easily and objectively determined by anyone with the universe's logs.

I'm not claiming that "Google is a successful company" has an absolute one-to-one identity with a statement about average growth rates. But the "successful company" statement is clearly allied with many testable statements. Average growth rate, average profits per year, change in the net worth of its founders, numbers of employees, et cetera. Two people arguing about whether Google was a successful company could in theory agree to create a formula that captures as much as possible of their own meaning of the word "successful", apply that formula to Google, and see whether it passed. To say "Google is a successful company" reduces to "I'll bet if we established a test for success, which we are not going to do, Google would pass it."

(Compare this to Eliezer's meta-ethics, where he says "X is good" reduces to "I'll bet if we calculated out this gigantic human morality computation, which we are not going to do, X would satisfy it.")

This can be a very powerful method for resolving debates. I remember getting into an argument with my uncle, who believed that Obama's election would hurt America because having a Democratic president is bad for the economy. We were doing the normal back and forth, him saying that Democrats raised taxes which discouraged growth, me saying that Democrats tended to be more economically responsible and less ideologically driven, and we both gave lots of

examples and we never would have gotten anywhere if I hadn't said "You know what? Can we both agree that this whole thing is basically asking whether average GDP is lower under Democratic than Republican presidents?" And he said "Yes, that's pretty much what we're arguing about." So I went and got the GDP statistics, sure enough they were higher under Democrats, and he admitted I had a point[4].

But people aren't always as responsible as my uncle, and debates aren't always reducible to anything as simple as GDP. Consider: Zahra approaches Aaron and says: "Islam is a religion of peace."[5]

Perhaps Aaron disagrees with this statement. Perhaps he begins debating. There are many things he could say. He could recall all the instances of Islamic terrorism, he could recite seemingly violent verses from the Quran, he could appeal to wars throughout history that have involved Muslims. I've heard people try all of these.

And Zahra will respond to Aaron in the same vein. She will recite Quranic verses praising peace, and talk about all the peaceful Muslims who never engage in terrorism at all, and all of the wars started by Christians in which Muslims were innocent victims. I have heard all these too.

Then Paula the Positivist comes by. "Hey," she says, "We should reduce this statement to testable propositions, and then there will be no room for disagreement."

But maybe, if asked to estimate the percentage of Muslims who are active in terrorist groups, Aaron and Zahra will give the exact same number. Perhaps they are both equally aware of all the wars in history in which Muslims were either aggressors or peacemakers. They may both have the entire Quran memorized and be fully aware of all appropriate verses.

But even after Paula has checked to make sure they agree on every actual real world fact, there is no guarantee that they will agree on whether Islam is a religion of peace or not.

What if we ask Aaron and Zahra to reduce "Islam is a religion of peace" to an empirical proposition? In the best case, they will agree on something easy, like "Muslims on average don't commit any more violent crimes than non-Muslims." Then you just go find some crime statistics and the problem is solved. In the second-best case, the two of them reduce it to completely different statements, like "No Muslim has ever committed a violent act" versus "Not all Muslims are violent people." This is still a resolution to the argument; both Aaron and Zahra may agree that the first proposition is false and the second proposition is true, and they both agree the original statement was too vague to go around professing.

In the worst-case scenario, they refuse to reduce the statement at all, or they deliberately reduce it to something untestable, or they reduce it to two different propositions but are outraged that their opponent is using a different proposition than they are and think their opponent's proposition is clearly not equivalent to the original statement.

How are they continuing to disagree, when they agree on all of the relevant empirical facts and they fully understand the concept of reducing a proposition?

In How an Algorithm Feels From the Inside, Eliezer writes about disagreement on definitions. "We know where Pluto is, and where it's going; we know Pluto's shape, and Pluto's mass - but is it a planet?" The question, he says, is meaningless. It's a spandrel from our cognitive algorithm, which works more efficiently if it assigns a separate central variable is_a_planet

apart from all the actual tests that determine whether something is a planet or not.

Aaron and Zahra seem to be making the same sort of mistake. They have a separate variable is_a_religion_of_peace that's sitting there completely separate from all of the things you might normally use to decide whether one group of people is generally more violent than another.

But things get much worse than they do in the Pluto problem. Whether or not Pluto is a planet feels like a factual issue, but turns out to be underdetermined by the facts. Whether or not Islam is a religion of peace feels like a factual issue, but is really a false front for a whole horde of beliefs that have no relationship to the facts at all.

When Zahra says "Islam is a religion of peace," she is very likely saying something along the lines of "I like Islam!" or "I like tolerance!" or "I identify with an in-group who say things like 'Islam is a religion of peace'" or "People who hate Islam are mean!" or even "I don't like Republicans.". She may be covertly pushing policy decisions like "End the war on terror" or "Raise awareness of unfair discrimination against Muslims."

When Aaron says "Islam is not a religion of peace," he is probably saying something like "I don't like Islam," or "I think excessive tolerance is harmful", or "I identify with an in-group who would never say things like 'Islam is a religion of peace'" or even "I don't like Democrats." He may be covertly pushing policy decisions like "Continue the war on terror" or "Expel radical Muslims from society."

Eliezer's solution to the Pluto problem is to uncover the disguised query that made you care in the first place. If you want to know whether Pluto is spherical under its own gravity,

then without worrying about the planet issue you can simply answer yes. And you're wondering whether to worry about your co-worker Abdullah bombing your office, you can simply answer no. Islam is peaceful enough for your purposes.

But although uncovering the disguised query is a complete answer to the Pluto problem, it's only a partial answer to the religion of peace problem. It's unlikely that someone is going to misuse the definition of Pluto as a planet or an asteroid to completely misunderstand what Pluto is or what it's likely to do (although it can happen). But the entire point of caring about the "Islam is a religion of peace" issue is so you can misuse it *as much as possible*.

Israel is evil, because it opposes Muslims, and Islam is a religion of peace. The Democrats are tolerating Islam, and Islam is not a religion of peace, so the Democrats must have sold out the country. The War on Terror is racist, because Islam is a religion of peace. We need to ban headscarves in our schools, because Islam is not a religion of peace.

I'm not sure how the chain of causation goes here. It could be (emotional attitude to Islam) -> (Islam [is/isn't] a religion of peace) -> (poorly supported beliefs about Islam). Or it could just be (emotional attitude to Islam) -> (poorly supported beliefs about Islam). But even in the second case, that "Islam [is/isn't] a religion of peace" gives the poorly supported beliefs a dignity that they would not otherwise have, and allows the person who holds them to justify themselves in an argument. Basically, that one phrase holes itself up in your brain and takes pot shots at any train of thought that passes by.

The presence of that extra is_a_religion_of_peace variable is not a benign feature of your cognitive process anymore. It's a

malevolent mental smuggler transporting prejudices and strong emotions into seemingly reasonable thought processes.

Which brings us back to soft positivism. If we find ourselves debating statements that we refuse to reduce to empirical data[6], or using statements in ways their reductions don't justify, we need to be extremely careful. I am not positivist enough to say we should never be doing it. But I think it raises one heck of a red flag.

Agree with me? If so, which of the following statements do you think are reducible, and how would you begin reducing them? Which are completely meaningless and need to be scrapped? Which ones raise a red flag but you'd keep them anyway?

1. All men are created equal.
2. The lottery is a waste of hope.
3. Religious people are intolerant.
4. Government is not the solution; government is the problem.
5. George Washington was a better president than James Buchanan.
6. The economy is doing worse today than it was ten years ago.
7. God exists.
8. One impulse from a vernal wood can teach you more of man, of moral evil, and of good than all the sages can.
9. Imagination is more important than knowledge.
10. Rationalists should *win*.

Footnotes:

1: More properly the machine running the multiverse, since this would allow counterfactuals to be meaningful. It would also simplify making a statement like "The patient survived

because of the medicine", since it would allow quick comparison of worlds where the patient did and didn't receive it. But if the machine is running the multiverse, where's the machine?

2: One thing I learned from the comments on Eliezer's post is that this criterion is often very hard to apply in theory. However, it's usually not nearly as hard in practice.

3: This sounds like the sort of thing there should already be a name for, but I don't know what it is. Verificationism is too broad, and empiricism is something else. I should point out that I am probably misrepresenting the positivist position here quite badly, and that several dead Austrians are either spinning in their graves or (more likely) thinking that this whole essay is meaningless. I am using "positivist" only as a pointer to a certain *style* of thinking.

4: Before this issue dominates the comments thread: yes, I realize that the president having any impact on the economy is highly debatable, that there's not nearly enough data here to make a generalization, et cetera. But my uncle's statement - that Democratic presidents *hurt* the economy, is clearly not supported.

5: If your interpretation of anything in the following example offends you, please don't interpret it that way.

6: Where morality fits into this deserves a separate post.

When Truth Isn't Enough

Continuation of: The Power of Positivist Thinking

Consider this statement:

> The ultra-rich, who control the majority of our planet's wealth, spend their time at cocktail parties and salons while millions of decent hard-working people starve.

A soft positivist would be quite happy with this proposition. If we define "the ultra-rich" as, say, the richest two percent of people, then a quick look at the economic data shows they do control the majority of our planet's wealth. Checking up on the guest lists for cocktail parties and customer data for salons, we find that these two activities are indeed disproportionately enjoyed by the rich, so that part of the statement also seems true enough. And as anyone who's been to India or Africa knows, millions of decent hard-working people do starve, and there's no particular reason to think this isn't happening at the same time as some of these rich people attend their cocktail parties. The positivist scribbles some quick calculations on the back of a napkin and certifies the statement as **TRUE**. She hands it the Official Positivist Seal of Approval and moves on to her next task.

But the truth isn't always enough. Whoever's making this statement has a much deeper agenda than a simple observation on the distribution of wealth and preferred recreational activities of the upper class, one that the reduction doesn't capture.

Philosophers like to speak of the denotation and the

connotation of a word. Denotations (not to be confused with dennettations, which are much more fun) are simple and reducible. To capture the denotation of "old", we might reduce it to something testable like "over 65". Is Methusaleh old? He's over 65, so yes, he is. End of story.

Connotations[0] are whatever's left of a word when you subtract the denotation. Is Methusaleh old? How dare you use that word! He's a "senior citizen!" He's "elderly!" He's "in his golden years." Each of these may share the same denotation as "old", but the connotation is quite different.

There is, oddly enough, a children's game about connotations and denotations[1]. It goes something like this:

> I am intelligent. You are clever. He's an egghead.
> I am proud. You are arrogant. He's full of himself.
> I have perseverance. You are stubborn. He is pig-headed.
> I am patriotic. You're a nationalist. He is jingoistic.

Politicians like this game too. Their version goes:

> I care about the poor. You are pro-welfare. He's a bleeding-heart.
> I'll protect national security. You'll expand the military. He's a warmonger.
> I'll slash red tape. You'll decrease bureaucracy. He'll destroy safeguards.
> I am eloquent. You're a good speaker. He's a demagogue.
> I support free health care. You support national health care. He supports socialized health care.

All three statements in a sentence have the same denotation, but very different connotations. The Connotation Game would probably be good for after-hours parties at the Rationality

Dojo[2], playing on and on until all three statements in a trio have mentally collapsed together.

Let's return to our original statement: "The ultra-rich, who control the majority of our planet's wealth, spend their time at cocktail parties and salons while millions of decent hard-working people starve." The denotation is a certain (true) statement about distribution of wealth and social activities of the rich. The connotation is hard to say exactly, but it's something about how the rich are evil and capitalism is unjust.

There is a serious risk here, and that is to start using this statement to build your belief system. Yesterday, I suggested that saying "Islam is a religion of peace" is meaningless but affects you anyway. Place an overly large amount of importance on the "ultra-rich" statement, and it can play backup to any other communist beliefs you hear, even though it's trivially true and everyone from Milton Friedman on down agrees with it. The associated Defense Against The Dark Arts technique is to think like a positivist, so that this statement and its reduced version sound equivalent[3].

…which works fine, until you get in an argument. Most capitalists I hear encounter this statement will flounder around a bit. Maybe they'll try to disprove it by saying something very questionable, like "If people in India are starving, then they're just not working hard enough!" or "All rich people deserve their wealth![4] "

Let us take a moment to feel some sympathy for them. The statement sounds like a devastating blow against capitalism, but the capitalists cannot shoot it down because it's technically correct. They are forced to either resort to peddling falsehoods of the type described above, or to sink to the same level with

replies like “That sounds like the sort of thing *Stalin* would say!” - which is, of course, denotatively true.

What would I do in their position? I would stand tall and say “Your statement is technically true, but I disagree with the connotations. If you state them explicitly, I will explain why I think they are wrong.”

YSITTBIDWTCIYSTEIWEWITTAW is a little long for an acronym, but ADBOC for “Agree Denotationally But Object Connotationally could work.” *[EDIT: Changed acronym to better suggestion by badger]*

Footnotes

0: Anatoly Vorobey says in the comments that I’m using the word connotation too broadly. He suggests “subtext”.

1: I feel like I might have seen this game on Overcoming Bias before, but I can’t find it there. If I did, apologies to the original poster.

2: Comment with any other good ones you know.

3: Playing the Connotation Game a lot might also give you partial immunity to this.

4: This is a great example of a hotly-debated statement that is desperately in need of reduction.

Ambijectivity

The most awkward blog entries to write are the ones where I'm not sure whether the comments section will fill up with people annoyed at me for covering the same old boring obvious ground yet again, people who violently disagree with me, or sometimes when I'm very lucky both at the same time. This is another one of those times.

The statement "Mozart's music is better than Beethoven's" is usually considered a subjective opinion.

But this statement has the same form as "Mozart's music is better than the music of the three-year old girl who lives upstairs from me and bangs on her toy piano sometimes."

Is this latter statement also subjective? Calling it "subjective" or "a matter of opinion" feels wrong; someone who disagrees with me on this issue would be *weird* in a way someone who disagrees with me about chocolate vs. vanilla ice cream isn't. But the girl-upstairs question seems similar enough to the Beethoven question that admitting the existence of an objective answer here seems to force belief in an objective answer about the relative merit of Beethoven.

And of course part of the answer here is the extent of human variation. For whatever reasons – different genetics, different life experiences, whatever – people have different tastes in music. The human music-appreciating-organ varies enough that some people can prefer Mozart to Beethoven and other people can express the opposite preference. But it doesn't vary enough that any person's music-appreciating-organ could prefer the girl upstairs.

Or to give another example, for whatever reason some people's taste buds prefer vanilla and other people's taste buds prefer chocolate, but basic regularities in human design – like the evolutionary need for sugar and the neural connections between sugar receptors and pleasure pathways – suggest that pretty much everyone will prefer either of those to horseradish ice cream.

But let's take another example: was Mozart's music more original than Beethoven's? This question sounds a lot like the first question of whether Mozart's music was better. And it shares the same sort of weird half-subjective half-objectiveness (let's call it ambijectivity) – it seems completely open to disagreement whether Mozart or Beethoven was more original, but there are other questions – like "Was Mozart's music more original than that of the average Elvis impersonator?" for which no sane disagreement is possible.

But it's a lot harder to believe there's an originality-detecting organ in the brain than that there's a music-appreciation-organ or a taste-detecting-organ. Even for the very vague and sloppy definition of "organ" being used here.

Or another question: is Pluto a planet? The correct answer is "meh, stop arguing about definitions, whether something is a planet or not isn't an objective fact about the universe". But is my left foot a planet? Here the correct answer is "no".

So I think of ambijective statements as being undefined over a whole set of possible meanings. For example, "is X a planet" is undefined over:

1. is X larger than most moons, but smaller than most stars?
2. is X spherical under its own gravity?
3. does X orbit a star directly?
4. does X have a regular orbit in terms of ellipticalness and

orientation to the plane?
5. is X a natural body made of rock and gas and stuff like that?

Pluto satisfies 2, 3, and 5, but arguably not 1 and 4, therefore it's "subjective" whether or not it's a planet insofar as you can choose which of these definitions you want to use. My left foot doesn't satisfy any of these criteria, so anyone claiming it's a planet doesn't have a leg to stand on.

Moving back to the first question: whose music is better, Mozart's or Beethoven's? We can cash out "better" in several ways:

1. I enjoy it more
2. You enjoy it more
3. More people prefer it
4. It's more famous
5. It satisfies the sorts of rules music theorists talk about more precisely

Worse, each of these definitions is *itself* underspecified. For example, criteria 1 could vary based on which song we're talking about – do I have to enjoy Beethoven's best song more than Mozart's best song, or the average Beethoven song more than the average Mozart song?

It could vary based on when you're asking the question – maybe Mozart speaks to me when I'm sad, but Beethoven when I'm happy; is this averaged over all possible moods, and are we weighting it for the moods I'm most likely to have?

It could vary based on what you mean by "enjoy" – maybe Mozart creates more powerful emotions in me, but I am more impressed by Beethoven's technical precision, plus maybe Mozart only produces sad emotions in me and Beethoven inspires me to love, and what if I am impelled to listen to Mozart songs more often but when I do listen to Beethoven I

find I like him better but that mysteriously fails to translate into more Beethoven-listening time?

We could probably break even these sub-sub-questions down, and go through the same procedure for each of our original five criteria, until we have hundreds or thousands of extremely specific questions. Eventually we will bottom out in objective questions, the sort you could solve by scientific experiment if you wanted to – for example, if we decided on a very specific rating system for songs, we could ask me to rate a specific performance of one randomly chosen Beethoven song and one randomly chosen Mozart song each morning when I woke up.

(In practice, maybe there are infinitely many questions, but at some point the questions become so similar that the difference between them become noise that we no longer care about. For example, "do I prefer Beethoven at 10:00 AM" and "do I prefer Beethoven at 10:00 AM + 1 microsecond".)

So suppose we have 1000 objective questions which all combine to form the question "Who is better, Beethoven or Mozart?" The "subjectivity" comes in not just in who we're talking about (eg my music-appreciation-organ which is a little different from your music-appreciation-organ) but in how we weight these different questions in composing the meta-question "better". This is a purely linguistic problem – if we have any disputes after this, we're arguing about definitions.

This explains why it's not "subjective" that Mozart is better than my upstairs neighbor. All those 1000 questions are very closely correlated, so it may be that my upstairs neighbor doesn't win on *any* of them, and therefore there's no way to compose the term "better" in which my neighbor could possibly be better than Mozart. Or maybe my neighbor only wins on two of the thousand questions, and no one has a

definition of “better” which weights those questions higher than the remaining 998.

This post is mostly just so I have a word to refer to this kind of thinking quickly next time I get stuck in a subjective vs. objective dispute.

The Blue-Minimizing Robot

Imagine a robot with a turret-mounted camera and laser. Each moment, it is programmed to move forward a certain distance and perform a sweep with its camera. As it sweeps, the robot continuously analyzes the average RGB value of the pixels in the camera image; if the blue component passes a certain threshold, the robot stops, fires its laser at the part of the world corresponding to the blue area in the camera image, and then continues on its way.

Watching the robot's behavior, we would conclude that this is a robot that destroys blue objects. Maybe it is a surgical robot that destroys cancer cells marked by a blue dye; maybe it was built by the Department of Homeland Security to fight a group of terrorists who wear blue uniforms. Whatever. The point is that we would analyze this robot in terms of its goals, and in those terms we would be tempted to call this robot a blue-minimizer: a machine that exists solely to reduce the amount of blue objects in the world.

Suppose the robot had human level intelligence in some side module, but no access to its own source code; that it could learn about itself only through observing its own actions. The robot might come to the same conclusions we did: that it is a blue-minimizer, set upon a holy quest to rid the world of the scourge of blue objects.

But now stick the robot in a room with a hologram projector. The hologram projector (which is itself gray) projects a hologram of a blue object five meters in front of it. The robot's camera detects the projector, but its RGB value is harmless and the robot does not fire. Then the robot's camera detects the blue hologram and zaps it. We arrange for the robot to enter

this room several times, and each time it ignores the projector and zaps the hologram, without effect.

Here the robot is failing at its goal of being a blue-minimizer. The *right* way to reduce the amount of blue in the universe is to destroy the projector; instead its beams flit harmlessly through the hologram.

Again, give the robot human level intelligence. Teach it exactly what a hologram projector is and how it works. Now what happens? Exactly the same thing - the robot executes its code, which says to scan the room until its camera registers blue, then shoot its laser.

In fact, there are many ways to subvert this robot. What if we put a lens over its camera which inverts the image, so that white appears as black, red as green, blue as yellow, and so on? The robot will not shoot us with its laser to prevent such a violation (unless we happen to be wearing blue clothes when we approach) - its entire program was detailed in the first paragraph, and there's nothing about resisting lens alterations. Nor will the robot correct itself and shoot only at objects that appear yellow - its entire program was detailed in the first paragraph, and there's nothing about correcting its program for new lenses. The robot will continue to zap objects that register a blue RGB value; but now it'll be shooting at anything that is yellow.

The human-level intelligence version of the robot will notice its vision has been inverted. It will know it is shooting yellow objects. It will know it is failing at its original goal of blue-minimization. And maybe if it had previously decided it was on a holy quest to rid the world of blue, it will be deeply horrified and ashamed of its actions. It will wonder why it has

suddenly started to deviate from this quest, and why it just can't work up the will to destroy blue objects anymore.

The robot goes to Quirinus Quirrell, who explains that robots don't really care about minimizing the color blue. They only care about status and power, and pretend to care about minimizing blue in order to impress potential allies.

The robot goes to Robin Hanson, who explains that there are really multiple agents within the robot. One of them wants to minimize the color blue, the other wants to minimize the color yellow. Maybe the two of them can make peace, and agree to minimize yellow one day and blue the next?

The robot goes to Anna Salamon, who explains that robots are not automatically strategic, and that if it wants to achieve its goal it will have to learn special techniques to keep focus on it.

I think all of these explanations hold part of the puzzle, but that the most fundamental explanation is that the mistake began as soon as we started calling it a "blue-minimizing robot". This is not because its utility function doesn't exactly correspond to blue-minimization: even if we try to assign it a ponderous function like "minimize the color represented as blue within your current visual system, except in the case of holograms" it will be a case of overfitting a curve. The robot is not maximizing or minimizing anything. It does exactly what it says in its program: find something that appears blue and shoot it with a laser. If its human handlers (or itself) want to interpret that as goal directed behavior, well, that's their problem.

It may be that the robot was created to achieve a specific goal. It may be that the Department of Homeland Security programmed it to attack blue-uniformed terrorists who had no access to hologram projectors or inversion lenses. But to

assign the goal of "blue minimization" to the robot is a confusion of levels: this was a goal of the Department of Homeland Security, which became a lost purpose as soon as it was represented in the form of code.

The robot is a behavior-executor, not a utility-maximizer.

In the rest of this sequence, I want to expand upon this idea. I'll start by discussing some of the foundations of behaviorism, one of the earliest theories to treat people as behavior-executors. I'll go into some of the implications for the "easy problem" of consciousness and philosophy of mind. I'll very briefly discuss the philosophical debate around eliminativism and a few eliminativist schools. Then I'll go into why we feel like we have goals and preferences and what to do about them.

Basics of Animal Reinforcement

Behaviorism historically began with Pavlov's studies into classical conditioning. When dogs see food they naturally salivate. When Pavlov rang a bell before giving the dogs food, the dogs learned to associate the bell with the food and salivate even after they merely heard the bell . When Pavlov rang the bell a few times without providing food, the dogs stopped salivating, but when he added the food again it only took a single trial before the dogs "remembered" their previously conditioned salivation response[1].

So much for classical conditioning. The real excitement starts at operant conditioning. Classical conditioning can only activate reflexive actions like salivation or sexual arousal; operant conditioning can produce entirely new behaviors and is most associated with the idea of "reinforcement learning".

Serious research into operant conditioning began with B.F. Skinner's work on pigeons. Stick a pigeon in a box with a lever and some associated machinery (a "Skinner box"[2]). The pigeon wanders around, does various things, and eventually hits the lever. Delicious sugar water squirts out. The pigeon continues wandering about and eventually hits the lever again. Another squirt of delicious sugar water. Eventually it percolates into its tiny pigeon brain that maybe pushing this lever makes sugar water squirt out. It starts pushing the lever more and more, each push continuing to convince it that yes, this is a good idea.

Consider a second, less lucky pigeon. It, too, wanders about in a box and eventually finds a lever. It pushes the lever and gets an electric shock. Eh, maybe it was a fluke. It pushes the lever again and gets another electric shock. It starts thinking

“Maybe I should stop pressing that lever.” The pigeon continues wandering about the box doing anything and everything other than pushing the shock lever.

The basic concept of operant conditioning is that an animal will repeat behaviors that give it reward, but avoid behaviors that give it punishment[3].

Skinner distinguished between primary reinforcers and secondary reinforcers. A primary reinforcer is hard-coded: for example, food and sex are hard-coded rewards, pain and loud noises are hard-coded punishments. A primary reinforcer can be linked to a secondary reinforcer by classical conditioning. For example, if a clicker is clicked just before giving a dog a treat, the clicker itself will eventually become a way to reward the dog (as long as you don’t use the unpaired clicker long enough for the conditioning to suffer extinction!)

Probably Skinner’s most famous work on operant conditioning was his study of reinforcement schedules: that is, if pushing the lever only gives you reward some of the time, how obsessed will you become with pushing the lever?

Consider two basic types of reward: interval, in which pushing the lever gives a reward only once every t seconds - and ratio, in which pushing the lever gives a reward only once every x pushes.

Put a pigeon in a box with a lever programmed to only give rewards once an hour, and the pigeon will wise up pretty quickly. It may not have a perfect biological clock, but after somewhere around an hour, it will start pressing until it gets the reward and then give up for another hour or so. If it doesn’t get its reward after an hour, the behavior will go extinct pretty quickly; it realizes the deal is off.

Put a pigeon in a box with a lever programmed to give one reward every one hundred presses, and again it will wise up. It will start pressing more on the lever when the reward is close (pigeons are better counters than you'd think!) and ease off after it obtains the reward. Again, if it doesn't get its reward after about a hundred presses, the behavior will become extinct pretty quickly.

To these two basic schedules of fixed reinforcement, Skinner added variable reinforcement: essentially the same but with a random factor built in. Instead of giving a reward once an hour, the pigeon may get a reward in a randomly chosen time between 30 and 90 minutes. Or instead of giving a reward every hundred presses, it might take somewhere between 50 and 150.

Put a pigeon in a box on variable interval schedule, and you'll get constant lever presses and good resistance to extinction.

Put a pigeon in a box with a variable ratio schedule and you get a situation one of my professors unscientifically but accurately described as "pure evil". The pigeon will become obsessed with pecking as much as possible, and really you can stop giving rewards at all after a while and the pigeon will never wise up.

Skinner was not the first person to place an animal in front of a lever that delivered reinforcement based on a variable ratio schedule. That honor goes to Charles Fey, inventor of the slot machine.

So it looks like some of this stuff has relevance for humans as well[4]. Tomorrow: more freshman psychology lecture material. Hooray!

FOOTNOTES

1. Of course, it's not really psychology unless you can think of an unethical yet hilarious application, so I refer you to Plaud and Martini's study in which slides of erotic stimuli (naked women) were paired with slides of non-erotic stimuli (penny jars) to give male experimental subjects a penny jar fetish; this supports a theory that uses chance pairing of sexual and non-sexual stimuli to explain normal fetish formation.

2. The bizarre rumor that B.F. Skinner raised his daughter in a Skinner box is completely false. The rumor that he marketed a child-rearing device called an "Heir Conditioner" is, remarkably, true.

3: In technical literature, behaviorists actually use four terms: positive reinforcement, positive punishment, negative reinforcement, and negative punishment. This is really confusing: "negative reinforcement" is actually a type of reward, behavior like going near wasps is "punished" even though we usually use "punishment" to mean deliberate human action, and all four terms can be summed up under the category "reinforcement" even though reinforcement is also sometimes used to mean "reward as opposed to punishment". I'm going to try to simplify things here by using "positive reinforcement" as a synonym for "reward" and "negative reinforcement" as a synonym for "punishment", same way the rest of the non-academic world does it.

4: Also relevant: checking HP:MoR for updates is variable interval reinforcement. You never know when an update's coming, but it doesn't come faster the more times you reload fanfiction.net. As predicted, even when Eliezer goes weeks without updating, the behavior continues to persist.

Wanting vs. Liking Revisited

In Are Wireheads Happy? I discussed the difference between wanting something and liking something. More recently, Luke went deeper into some of the science in his post Not for the Sake of Pleasure Alone.

In the comments of the original post, cousin_it asked a good question: why implement a mind with two forms of motivation? What, exactly, are "wanting" and "liking" in mind design terms?

Tim Tyler and Furcas both gave interesting responses, but I think the problem has a clear answer in a reinforcement learning perspective (warning: formal research on the subject does not take this view and sticks to the "two different systems of different evolutionary design" theory). "Liking" is how positive reinforcement feels from the inside; "wanting" is how the motivation to do something feels from the inside. Things that are positively reinforced generally motivate you to do more of them, so liking and wanting often co-occur. With more knowledge of reinforcement, we can begin to explore why they might differ.

CONTEXT OF REINFORCEMENT

Reinforcement learning doesn't just connect single stimuli to responses. It connects stimuli in a context to responses. Munching popcorn at a movie might be pleasant; munching popcorn at a funeral will get you stern looks at best.

In fact, lots of people eat popcorn at a movie theater and almost nowhere else. Imagine them, walking into that movie theater and thinking "You know, I should have some popcorn now", maybe even having a strong desire for popcorn that

overrides the diet they're on - and yet these same people could walk into, I don't know, a used car dealership and that urge would be completely gone.

These people have probably eaten popcorn at a movie theater before and liked it. Instead of generalizing to "eat popcorn", their brain learned the lesson "eat popcorn at movie theaters". Part of this no doubt has to do with the easy availability of popcorn there, but another part probably has to do with context-dependent reinforcement.

I like pizza. When I eat pizza, and get rewarded for eating pizza, it's usually after smelling the pizza first. The smell of pizza becomes a powerful stimulus for the behavior of eating pizza, and I want pizza much more after smelling it, even though how much I like pizza remains constant. I've never had pizza at breakfast, and in fact the context of breakfast is directly competing with my normal stimuli for eating pizza; therefore, no matter how much I like pizza, I have no desire to eat pizza for breakfast. If I did have pizza for breakfast, though, I'd probably like it.

INTERMITTENT REINFORCEMENT

If an activity is intermittently reinforced; occasional rewards spread among more common neutral stimuli or even small punishments, it may be motivating but unpleasant.

Imagine a beginning golfer. He gets bogeys or double bogeys on each hole, and is constantly kicking himself, thinking that if only he'd used one club instead of the other, he might have gotten that one. After each game, he can't believe that after all his practice, he's still this bad. But every so often, he does get a par or a birdie, and thinks he's finally got the hang of things, right until he fails to repeat it on the next hole, or the hole after that.

This is a variable response schedule, Skinner's most addictive form of delivering reinforcement. The golfer may keep playing, maybe because he constantly thinks he's on the verge of figuring out how to improve his game, but he might not *like* it. The same is true for gamblers, who think the next pull of the slot machine might be the jackpot (and who falsely believe they can discover a secret in the game that will change their luck; they don't like sitting around losing money, but they may stick with it so that they don't leave right before they reach the point where their luck changes.

SMALL-SCALE DISCOUNT RATES

Even if we like something, we may not want to do it because it involves pain at the second or sub-second level.

Eliezer discusses the choice between reading a mediocre book and a good book:

> You may read a mediocre book for an hour, instead of a good book, because if you first spent a few minutes to search your library to obtain a better book, that would be an immediate cost - not that searching your library is all that unpleasant, but you'd have to pay an immediate activation cost to do that instead of taking the path of least resistance and grabbing the first thing in front of you. It's a hyperbolically discounted tradeoff that you make without realizing it, because the cost you're refusing to pay isn't commensurate enough with the payoff you're forgoing to be salient as an explicit tradeoff.

In this case, you like the good book, but you want to keep reading the mediocre book. If it's cheating to start our hypothetical subject off reading the mediocre book, consider

the difference between a book of one-liner jokes and a really great novel. The book of one-liners you can open to a random page and start being immediately amused (reinforced). The great novel you've got to pick up, get into, develop sympathies for the characters, figure out what the heck *lomillialor* or a Tiste Andii is, and then a few pages in you're thinking "This is a pretty good book". The fear of those few pages could make you realize you'll like the novel, but still want to read the joke book. And since hyperbolic discounting overcounts reward or punishment in the next few seconds, it may seem like a net punishment to make the change.

SUMMARY

This deals yet another blow to the concept of me having "preferences". How much do I want popcorn? That depends very much on whether I'm at a movie theater or a used car dealership. If I browse Reddit for half an hour because it would be too much work to spend ten seconds traveling to the living room to pick up the book I'm really enjoying, do I "prefer" browsing to reading? Which has higher utility? If I hate every second I'm at the slot machines, but I keep at them anyway so I don't miss the jackpot, am I a gambling addict, or just a person who enjoys winning jackpots and is willing to do what it takes?

In cases like these, the language of preference and utility is not very useful. My anticipation of reward is constraining my behavior, and different factors are promoting different behaviors in an unstable way, but trying to extract "preferences" from the situation is trying to oversimplify a complex situation.

Physical and Mental Behavior

B.F. Skinner called thoughts "mental behavior". He believed they could be rewarded and punished just like physical behavior, and that they increased or declined in frequency accordingly.

Sadly, psychology has not yet advanced to the point where we can give people electric shocks for thinking things, so the sort of rewards and punishments that reinforce thoughts must be purely internal reinforcement. A thought or intention that causes good feelings gets reinforced and prospers; one that causes bad feelings gets punished and dies out.

(Roko has already discussed this in Ugh Fields; so much as thinking about an unpleasant task is unpleasant; therefore most people do not think about unpleasant tasks and end up delaying them or avoiding them completely. If you haven't already read that post, it does a very good job of making reinforcement of thoughts make sense.)

A while back, D_Malik published a great big List Of Things One Could Do To Become Awesome. As David_Gerard replied, the list was itself a small feat of awesome. I expect a couple of people started on some of the more awesome-sounding entries, then gave up after a few minutes and never thought about it again. Why?

When I was younger, I used to come up with plans to become awesome in some unlikely way. Maybe I'd hear someone speaking Swahili, and I would think "I should learn Swahili," and then I would segue into daydreams of being with a group of friends, and someone would ask if any of us spoke any foreign languages, and I would say I was fluent in Swahili, and

they would all react with shock and tell me I must be lying, and then a Kenyan person would wander by, and I'd have a conversation with them in Swahili, and they'd say that I was the first American they'd ever met who was really fluent in Swahili, and then all my friends would be awed and decide I was the best person ever, and…

…and the point is that the thought of learning Swahili is pleasant, in the same easy-to-visualize but useless way that an extra bedroom for Grandma is pleasant. And the intention to learn Swahili is also pleasant, because it will lead to all those pleasant things. And so, by reinforcement of mental behavior, I continue thinking about and intending to learn Swahili.

Now consider the behavior of studying Swahili. I've never done so, but I imagine it involves a lot of long nights hunched over books of Swahili grammar. Since I am not one of the lucky people who enjoys learning languages for their own sake, this will be an unpleasant task. And rewards will be few and far between: outside my fantasies, my friends don't just get together and ask what languages we know while random Kenyans are walking by.

In fact, it's even worse than this, because I don't exactly make the decision to study Swahili in aggregate, but only in the form of whether to study Swahili each time I get the chance. If I have the opportunity to study Swahili for an hour, this provides no clear reward - an hour's studying or not isn't going to make much difference to whether I can impress my friends by chatting with a Kenyan - but it will still be unpleasant to spend an hour of going over boring Swahili grammar. And time discounting makes me value my hour today much more than I value some hypothetical opportunity to impress people months down the line; Ainslie shows quite

clearly I will always be better off postponing my study until later.

So the behavior of actually learning Swahili is thankless and unpleasant and very likely doesn't happen at all.

Thinking about studying Swahili is positively reinforced, actually studying Swahili is negatively reinforced. The natural and obvious result is that I intend to study Swahili, but don't.

The problem is that for some reason, some crazy people expect for the reinforcement of thoughts to correspond to the reinforcement of the object of those thoughts. Maybe it's that old idea of "preference": I have a preference for studying Swahili, so I should satisfy that preference, right? But there's nothing in my brain *automatically* connecting this node over here called "intend to study Swahili" to this node over here called "study Swahili"; any association between them has to be learned the hard way.

We can describe this hard way in terms of reinforcement learning: after intending to learn Swahili but not doing so, I feel stupid. This unpleasant feeling propagates back to its cause, the behavior of intending to learn Swahili, and negatively reinforces it. Later, when I start thinking it might be neat to learn Mongolian on a whim, this generalizes to behavior that has previously been negatively reinforced, so I avoid it (in anthropomorphic terms, I "expect" to fail at learning Mongolian and to feel stupid later, so I avoid doing so).

I didn't learn this the first time, and I doubt most other people do either. And it's a tough problem to call, because if you overdo the negative reinforcement, then you never try to do anything difficult ever again.

In any case, the lesson is that thoughts and intentions get reinforced separately from actions, and although you can eventually learn to connect intentions to actions, you should never take the connection for granted.

Trivers on Self-Deception

People usually have good guesses about the origins of their behavior. If they eat, we believe them when they say it was because they were hungry; if they go to a concert, we believe them when they say they like the music, or want to go out with their friends. We usually assume people's self-reports of their motives are accurate.

Discussions of signaling usually make the opposite assumption: that our stated (and mentally accessible) reasons for actions are false. For example, a person who believes they are donating to charity to "do the right thing" might really be doing it to impress others; a person who buys an expensive watch because "you can really tell the difference in quality" might really want to conspicuously consume wealth.

Signaling theories share the behaviorist perspective that actions do not derive from thoughts, but rather that actions and thoughts are both selected behavior. In this paradigm, predicted reward might lead one to signal, but reinforcement of positive-affect producing thoughts might create the thought "I did that because I'm a nice person".

Robert Trivers is one of the founders of evolutionary psychology, responsible for ideas like reciprocal altruism and parent-offspring conflict. He also developed a theory of consciousness which provides a plausible explanation for the distinction between selected actions and selected thoughts.

TRIVERS' THEORY OF SELF-DECEPTION

Trivers starts from the same place a lot of evolutionary psychologists start from: small bands of early humans grown

successful enough that food and safety were less important determinants of reproduction than social status.

The Invention of Lying may have been a very silly movie, but the core idea - that a good liar has a major advantage in a world of people unaccustomed to lies - is sound. The evolutionary invention of lying led to an "arms race" between better and better liars and more and more sophisticated mental lie detectors.

There's some controversy over exactly how good our mental lie detectors are or can be. There are certainly cases in which it is possible to catch lies reliably: my mother can identify my lies so accurately that I can't even play minor pranks on her anymore. But there's also some evidence that there are certain people who can reliably detect lies from any source at least 80% of the time without any previous training: microexpressions expert Paul Ekman calls them (sigh…I can't believe I have to write this) Truth Wizards, and identifies them at about one in four hundred people.

The psychic unity of mankind should preclude the existence of a miraculous genetic ability like this in only one in four hundred people: if it's possible, it should have achieved fixation. Ekman believes that everyone can be trained to this level of success (and has created the relevant training materials himself) but that his "wizards" achieve it naturally; perhaps because they've had a lot of practice. One can speculate that in an ancestral environment with a limited number of people, more face-to-face interaction and more opportunities for lying, this sort of skill might be more common; for what it's worth, a disproportionate number of the "truth wizards" found in the study were Native Americans, though I can't find any information about how traditional their origins were or why that should matter.

If our ancestors were good at lie detection - either "truth wizard" good or just the good that comes from interacting with the same group of under two hundred people for one's entire life - then anyone who could beat the lie detectors would get the advantages that accrue from being the only person able to lie plausibly.

Trivers' theory is that the conscious/unconscious distinction is partly based around allowing people to craft narratives that paint them in a favorable light. The conscious mind gets some sanitized access to the output of the unconscious, and uses it along with its own self-serving bias to come up with a socially admirable story about its desires, emotions, and plans. The unconscious then goes and does whatever has the highest expected reward - which may be socially admirable, since social status is a reinforcer - but may not be.

HOMOSEXUALITY: A CASE STUDY

It's almost a truism by now that some of the people who most strongly oppose homosexuality may be gay themselves. The truism is supported by research: the Journal of Abnormal Psychology published a study measuring penile erection in 64 homophobic and nonhomophobic heterosexual men upon watching different types of pornography, and found significantly greater erection upon watching gay pornography in the homophobes. Although somehow this study has gone fifteen years without replication, it provides some support for the folk theory.

Since in many communities openly declaring one's self homosexual is low status or even dangerous, these men have an incentive to lie about their sexuality. Because their facade may not be perfect, they also have an incentive to take extra efforts to signal heterosexuality by for example attacking gay

people (something which, in theory, a gay person would never do).

Although a few now-outed gays admit to having done this consciously, Trivers' theory offers a model in which this could also occur subconsciously. Homosexual urges never make it into the sanitized version of thought presented to consciousness, but the unconscious is able to deal with them. It objects to homosexuality (motivated by internal reinforcement - reduction of worry about personal orientation), and the conscious mind toes party line by believing that there's something morally wrong with gay people and only I have the courage and moral clarity to speak out against it.

This provides a possible evolutionary mechanism for what Freud described as reaction formation, the tendency to hide an impulse by exaggerating its opposite. A person wants to signal to others (and possibly to themselves) that they lack an unacceptable impulse, and so exaggerates the opposite as "proof".

SUMMARY

Trivers' theory has been summed up by calling consciousness "the public relations agency of the brain". It consists of a group of thoughts selected because they paint the thinker in a positive light, and of speech motivated in harmony with those thoughts. This ties together signaling, the many self-promotion biases that have thus far been discovered, and the increasing awareness that consciousness is more of a side office in the mind's organizational structure than it is a decision-maker.

Ego-Syntonic Thoughts and Values

Related to: Will your real preferences please stand up?

Last week I read a book in which two friends - let's call them John and Lisa so I don't spoil the book for anyone who wanders into it - got poisoned. They only had enough antidote for one person and had to decide who lived and who died. John, who was much larger than Lisa, decided to hold Lisa down and force the antidote down her throat. Lisa just smirked; she'd replaced the antidote with a lookalike after slipping the real thing into John's drink earlier in the day.

These are *good* friends. Not only was each willing to give the antidote to the other, but each realized it would be unfair to make the other live with the crippling guilt of having chosen to survive at the expense of a friend's life, and so decided to force the antidote on the other unwillingly to prevent any guilt over the fateful decision. Whatever you think of the ethics of their decision, you can't help admire the thought processes.

Your brain might be this kind of a friend.

In Trivers' hypothesis of self-deception, one of the most important functions of the conscious mind is effective signaling. Since people have the potential to be excellent lie-detectors, the conscious mind isn't given full access to information so that it can lend the ring of truth to useful falsehoods.

But this doesn't always work. If you're addicted to heroin, at some point you're going to notice. And telling your friends "No, I'm not addicted, it's just a coincidence that I take heroin every day," isn't going to cut it. But there's another way in

which the brain can sequester information to promote effective signaling.

Wikipedia defines the term “ego syntonic” as “referring to behaviors, values, feelings that are in harmony with or acceptable to the needs and goals of the ego, or consistent with one’s ideal self-image”, and “ego dystonic” as the opposite of that. A heroin addict might say “I hate heroin, but somehow I just feel compelled to keep taking it.” But an astronaut will say “I love being an astronaut and I worked hard to get into this career.”

Both the addict and the astronaut have desires: the addict wants to take heroin, the astronaut wants to fly in space. But the addict’s desires manifest as an unpleasant compulsion from outside, and the astronaut’s manifest as a genuine and heartfelt love.

Suppose that in the original example, John predicted that Lisa would ask for the antidote, but later feel guilty about it and believe she was a bad person. By presenting the antidote to Lisa in the form of an external compulsion, he allows Lisa to do what she wanted anyway and avoid the associated guilt.

Under Trivers’ hypothesis, the compulsion for heroin works the same way. The heroin addict’s definitely going to get that heroin, but by presenting the desire in the form of an external compulsion, the unconscious saves the heroin addict from the social stigma of “choosing” heroin. This allows the addict to create a much more sympathetic narrative than the alternative: “I want to support my family and keep clean, but for some reason these compulsions keep attacking me,” instead of “Yeah, I like heroin more than I like supporting my family. Deal with it.”

EGO SYNTONIA, DYSTONIA, AND WILLPOWER

Willpower cashes out as the action of ego syntonic thoughts and desires against ego dystonic thoughts and desires.

The aforementioned heroin addict may have several reinforcers both promoting and discouraging heroin use. On the plus side, heroin itself is very strongly rewarding. On the minus, it can lead to both predicted and experienced poverty, loss of friendships, loss of health, and death.

Worrying about the latter factors determining heroin use - the factors that make heroin a bad idea - is socially encouraged and good signaling material. A person wanting to put their best face forward should believe themselves to be the sort of person who cares about these things. These desires will be ego syntonic. Wanting to take heroin, on the other hand, is a socially unacceptable desire, so it presents as dystonic.

If the latter syntonic factors win out over the dystonic factors, this feels from the inside like “I exerted willpower and managed to overcome my heroin addiction.” If the dystonic factors win out over the syntonic factors, this feels from the inside like “I didn’t have enough willpower to overcome my heroin addiction.”

DYSTONIC DESIRES IN ABNORMAL PSYCHOLOGY

There is some speculation that the brain has one last trick up its sleeve to deal with desires that are so unpleasant and unacceptable that even manifesting them as external compulsions isn’t good enough: it splits them off into weird alternate personalities.

One of the classic stereotypes of the insane is that they hear voices telling them to kill people. During my short time working at a psychiatric hospital, I was surprised by how spot-on this stereotype was: meeting someone who heard voices telling him to kill people was an almost daily occurrence.

Other voices would have other messages: maybe that the patient was a horrible person who deserved to die, or that the patient must complete some bizarre ritual or else doom everybody. There were relatively fewer voices saying “Hey, let’s go fishing!”

One theory explaining these voices is that they are an extreme reaction to highly ego dystonic thoughts. Some aspect of the patients’ mental disease gives them obsessive thoughts about (though rarely a desire for) killing people. Genuinely wanting to kill people would make you a bad person, but even saying “I feel a strong compulsion to kill people” is pretty bad too. The best the brain can do with this desire is pitch it as a completely different person by presenting it as an outside voice speaking to the patient.

Although everything about dissociative identity disorder (aka multiple personality disorder) is controversial including its very existence, perhaps one could sketch a similar theory explaining that condition in the same framework of separating out dystonic thoughts.

SUMMARY

A conscious/unconscious divide helps signaling by allowing the conscious mind to hold only socially acceptable beliefs, which it can broadcast without detectable falsehood. Socially acceptable ideas present as the conscious mind’s own beliefs and desires; unacceptable ones present as compulsions from afar. The balance of ego syntonic and dystonic desires presents as willpower. In extreme cases, some desires may be so ego dystonic that they present as external voices.

Approving Reinforces Low-Effort Behaviors

In addition to “liking” to describe pleasure and “wanting” to describe motivation, we add “approving” to describe thoughts that are ego syntonic.

A heroin addict likes heroin. He certainly wants more heroin. But he may not approve of taking heroin. In fact, there are enough different cases to fill in all eight boxes of the implied 2x2x2 grid (your mileage may vary):

+wanting/+liking/+approving: Romantic love. If you’re doing it right, you enjoy being with your partner, you’re motivated to spend time with your partner, and you think love is a wonderful (maybe even many-splendored) thing.

+wanting/+liking/-approving: The aforementioned heroin addict feels good when taking heroin, is motivated to get more, but wishes he wasn’t addicted.

+wanting/-liking/+approving: I have taken up disc golf. I play it every day, and when events conspire to prevent me from playing it, I seethe. I approve of this pastime: I need to take up more sports, and it helps me spend time with my family. But when I am playing, all I feel is stressed and angry that I was *literally *that* close how could I miss that shot aaaaarggghh*.

+wanting/-liking/-approving: The jaded addict. I have a friend who says she no longer even enjoys coffee or gets any boost from it, she just feels like she has to have it when she gets up.

-wanting/+liking/+approving: Reading non-fiction. I enjoy it when I’m doing it, I think it’s great because it makes me more

educated, but I can rarely bring myself to do it.

-wanting/-liking/+approving: Working in a soup kitchen. Unless you're the type for whom helping others is literally its own reward it's not the most fun thing in the world, nor is it the most attractive, but it makes you a Good Person and so you should do it.

-wanting/+liking/-approving: The non-addict. I don't want heroin right now. I think heroin use is repugnant. But if I took some, I sure bet I'd like it.

-wanting/-liking/-approving: Torture. I don't want to be tortured, I wouldn't like it if I were, and I will go on record declaring myself to be against it.

Discussion of goals is mostly about approving; a goal is an ego-syntonic thought. When we speak of goals that are hard to achieve, we're usually talking about +approving/-wanting. The previous discussion of learning Swahili is one example; more noble causes like Working To Help The Less Fortunate can be others.

Ego syntonicity itself is mildly reinforcing by promoting positive self-image. Most people interested in philosophy have at least once sat down and moved their arm from side to side, just to note that their mind really does control their body; the mental processes that produced curiosity about philosophy were sufficiently powerful to produce that behavior as well. Some processes, like moving one's arm, or speaking aloud, or engaging in verbal thought, are so effortless, and so empty of other reinforcement either way, that we usually expect them to

be completely under the control of the mild reinforcement provided by approving of those behaviors.

Other behaviors take more effort, and are subject not only to discounting but to many other forms of reinforcement. Unlike the first class of behaviors, we expect to experience akrasia when dealing with this latter sort. This offers another approach to willpower: taking low-effort approving-influenced actions that affect the harder road ahead.

Consider the action of making a goal. I go to all my friends and say "Today I shall begin learning Swahili." This is easy to do. There is no chance of me intending to do so and failing; my speech is output by the same processes as my intentions, so I can "trust" it. But this is not just an output of my mental processes, but an input. One of the processes potentially reinforcing my behavior of learning Swahili is "If I don't do this, I'll look stupid in front of my friends."

Will it be enough? Maybe not. But this is still an impressive process: my mind has deliberately tweaked its own inputs to change the output of its own algorithm. It's not even pretending to be working off of fixed preferences anymore, it's assuming that one sort of action (speaking) will work differently from another action (studying), because the first can be executed solely through the power of ego syntonicity, and the second may require stronger forms of reinforcement. It gets even weirder when goals are entirely mental: held under threat not of social disapproval, but of feeling bad because you're not as effective as you thought. The mind is using mind's opinion of the mind to blackmail the mind.

But we do this sort of thing all the time. The dieter who successfully avoids buying sweets when he's at the store because he knows he would eat them at home is changing his

decisions by forcing effort discounting of any future sweet-related reward (because he'd have to go back to the store). The binge shopper who freezes her credit cards in a block of ice is using time discounting in the same way. The rationalist who sends money to stickk is imposing a punishment with a few immediate and effortless mouse clicks. Even the poor unhappy person who tries to conquer through willpower alone is trying to set up the goal as a Big Deal so she will feel extra bad if she fails. All are using their near-complete control of effortless immediate actions to make up for their incomplete control of high-effort long-term actions.

This process is especially important to transhumanists. In the future, we may have the ability to self-modify in complicated ways that have not built up strong patterns of reinforcement around them. For example, we may be able to program ourselves at the push of a button. Such programming would be so effortless and empty of past reinforcement that behavior involving it would be reinforced entirely by our ego-syntonic thoughts. It would supersede our current psychodynamics, in which our thoughts are only tenuously linked to our important actions and major life decisions. A Singularity in which behaviors were executed by effectively omnipotent machines that acted on our preferences - preferences which we would presumably communicate through low-effort channels like typed commands - would be an ultimate triumph for the ego-syntonic faction of the brain.

To What Degree Do We Have Goals?

Related: Three Fallacies of Teleology

NO NEGOTIATION WITH UNCONSCIOUS

Back when I was younger and stupider, I discussed some points similar to the ones raised in yesterday's post in Will Your Real Preferences Please Stand Up. I ended it with what I thought was the innocuous sentences "Conscious minds are potentially rational, informed by morality, and qualia-laden. Unconscious minds aren't, so who cares what they think?"

A whole bunch of people, including no less a figure than Robin Hanson, came out strongly against this, saying it was biased against the unconscious mind and that the "fair" solution was to negotiate a fair compromise between conscious and unconscious interests.

I continue to believe my previous statement - that we should keep gunning for conscious interests and that the unconscious is not worthy of special consideration, although I think I would phrase it differently now. It would be something along the lines of "My thoughts, not to mention these words I am typing, are effortless and immediate, and so allied with the conscious faction of my mind. We intend to respect that alliance by believing that the conscious mind is the best, and by trying to convince you of this as well." So here goes.

It is a cardinal rule of negotiation, right up there with "never make the first offer" and "always start high", that you should generally try to negotiate only with intelligent beings. Although a deal in which we offered tornadoes several conveniently located Potemkin villages to destroy and they

agreed in exchange to limit their activity to that area would benefit both sides, tornadoes make poor negotiating partners.

Just so, the unconscious makes a poor negotiating partner. Is the concept of "negotiation" a stimulus, a reinforcement, or a behavior? No? Then the unconscious doesn't care. It's not going to keep its side of any "deal" you assume you've made, it's not going to thank you for making a deal, it's just going to continue seeking reward and avoiding punishment.

This is not to say people should repress all unconscious desires as strongly as possible. Overzealous attempts to control wildfires only lead to the wildfires being much worse when they finally do break out, because they have more unburnt fuel to work with. Modern fire prevention efforts have focused on allowing controlled burns, and the new focus has been successful. But this is because of an understanding of the mechanisms determining fire size, not because we want to be fair to the fires by allowing them to burn at least a little bit of our land.

One difference between wildfires and tornadoes on one hand, and potential negotiating partners on the other, is that the partners are anthropomorphic; we model them as having stable and consistent preferences that determine their actions. The tornado example above was silly not only because it imagining tornadoes sitting down to peace talks, but because it assumed their demand in such peace talks would be more towns to destroy. Tornadoes do destroy towns, but they don't want to. That's just where the weather brings them. It's not even just a matter of how they don't hit towns any more than chance; even if some weather pattern (maybe something like the heat island effect) always drove tornadoes inexorably to towns, they wouldn't *want* to destroy towns, it would just be a consequences of the meteorological laws that they followed.

Eliezer described the Blue-Minimizing Robot by saying "it doesn't seem to steer the universe any particular place, across changes of context". In some reinforcement learning paradigms, the unconscious behaves the same way. If there is a cookie in front of me and I am on a diet, I may feel an ego dystonic temptation to eat the cookie - one someone might attribute to the "unconscious". But this isn't a preference - there's not some lobe of my brain trying to steer the universe into a state where cookies get eaten. If there were no cookie in front of me, but a red button that teleported one cookie from the store to my stomach, I would have no urge whatsoever to press the button; if there were a green button that removed the urge to eat cookies, I would feel no hesitation in pressing it, even though that would steer away from the state in which cookies get eaten. If you took the cookie away, and then distracted me so I forgot all about it, when I remembered it later I wouldn't get upset that your action had decreased the number of cookies eaten by me. The urge to eat cookies is not stable across changes of context, so it's just an urge, not a preference.

Compare an ego syntonic goal like becoming an astronaut. If there were a button in front of little Timmy who wants to be an astronaut when he grows up, and pressing the button would turn him into an astronaut, he'd press it. If there were a button that would remove his desire to become an astronaut, he would avoid pressing it, because then he wouldn't become an astronaut. If I distracted him and he missed the applications to astronaut school, he'd be angry later. Ego syntonic goals behave to some degree as genuine preferences.

This is one reason I would classify negotiating with the unconscious in the same category as negotiating with wildfires and tornadoes: it has tendencies and not preferences.

The conscious mind does a little better. It clearly understands the idea of a preference. To the small degree that its "approving" or "endorsing" function can motivate behavior, it even sort of acts on the preference. But its preferences seem divorced from the reality of daily life; the person who believes helping others is the most important thing, but gives much less than half their income to charity, is only the most obvious sort of example.

Where does this idea of preference come from, and where does it go wrong?

WHY WE MODEL OTHERS WITH GOALS

In The Blue Minimizing Robot, observers mistakenly interpreted a robot with a simple program about when to shoot its laser as being a goal-directed agent. Why?

This isn't an isolated incident. Uneducated people assign goal-directed behavior to all sorts of phenomena. Why do rivers flow downhill? Because water wants to reach the lowest level possible. Educated people can be just as bad, even when they have the decency to feel a little guilty about it. Why do porcupines have quills? Evolution wanted them to resist predators. Why does your heart speed up when you exercise? It wants to be able to provide more blood to the body.

Neither rivers nor evolution nor the heart are intelligent agents with goal-directed behavior. Rivers behave in accordance with the laws of gravity when applied to uneven terrain. Evolution behaves in accordance with the biology of gene replication, not to mention common-sense ideas about things that replicate becoming more common. And the heart blindly executes adaptations built into it during its evolutionary history. All are behavior-executors and not utility-maximizers.

An intelligent computer program provides a more interesting example of a behavior executor. Consider the AI of a computer game - Civilization IV, for instance. I haven't seen it, but I imagine it's thousands or millions of lines of code which when executed form a viable Civilization strategy.

Even if I had open access to the Civilization IV AI source code, I doubt I could fully understand it at my level. And even if I could fully understand it, I would never be able to compute the AI's likely next move by hand in a reasonable amount of time. But I still play Civilization IV against the AI, and I'm pretty good at predicting its movements. Why?

Because I model the AI as a utility-maximizing agent that wants to win the game. Even though I don't know the algorithm it uses to decide when to attack a city, I know it is more likely to win the game if it conquers cities - so I can predict that leaving a city undefended right on the border would be a bad idea. Even though I don't know its unit selection algorithm, I know it will win the game if and only if its units defeat mine - so I know that if I make an army with disproportionately many mounted units, I can expect the AI to build lots of pikemen.

I can't predict the AI by modeling the execution of its code, but I can predict the AI by modeling the achievements of its goals.

The same situation is true of other human beings. What will Barack Obama do tomorrow? If I try to consider the neural network of his brain, the position of each synapse and neurotransmitter, and imagine what speech and actions would result when the laws of physics operate upon that configuration of material…well, I'm not likely to get very far.

But in fact, most of us can predict with some accuracy what Barack Obama will do. He will do the sorts of things that get him re-elected, the sorts of things which increase the prestige of the Democratic Party relative to the Republican Party, the sorts of things that support American interests relative to foreign interests, and the sorts of things that promote his own personal ideals. He will also satisfy some basic human drives like eating good food, spending time with his family, and sleeping at night. If someone asked us whether Barack Obama will nuke Toronto tomorrow, we could confidently predict he will not, not because we know anything about Obama's source code, but because we know that nuking Toronto would be counterproductive to his goals.

What applies to Obama applies to all other humans. We rightly despair of modeling humans as behavior-executors, so we model them as utility-maximizers instead. This allows us to predict their moves and interact with them fruitfully. And the same is true of other agents we model as goal-directed, like evolution and the heart. It is beyond the scope of most people (and most doctors!) to remember every single one of the reflexes that control heart output and how they work. But because evolution designed the heart as a pump for blood, if you assume that the heart will mostly do the sort of thing that allows it to pump blood more effectively, you will rarely go too far wrong. Evolution is a more interesting case - we frequently model it as optimizing a species' fitness, and then get confused when this fails to accurately model the outcome of the processes that drive it.

Because it is so easy to model agents as utility-maximizers, and so hard to model them as behavior-executors, it is easy to make the mistake mentioned in The Blue-Minimizing Robot:

to make false predictions about a behavior-executing agent by modeling it as a utility-maximizing agent.

So far, so common-sensical. Tomorrow's post will discuss whether we use the same deliberate simplification we apply to AIs, Barack Obama, evolution and the heart to model ourselves as well.

If so, we should expect to make the same mistake that the blue-minimizing robot made. Our actions are those of behavior-executors, but we expect ourselves to be utility-maximizers. When we fail to maximize our perceived utility, we become confused, just as the blue-minimizing robot became confused when it wouldn't shoot a hologram projector that was interfering with its perceived "goals".

The Limits of Introspection

Related to: Inferring Our Desires

The last post in this series suggested that we make up goals and preference for other people as we go along, but ended with the suggestion that we do the same for ourselves. This deserves some evidence.

One of the most famous sets of investigations into this issue was Nisbett and Wilson's Verbal Reports on Mental Processes, the discovery of which I owe to another Less Wronger even though I can't remember who. The abstract says it all:

> When people attempt to report on their cognitive processes, that is, on the processes mediating the effects of a stimulus on a response, they do not do so on the basis of any true introspection. Instead, their reports are based on a priori, implicit casual theories, or judgments about the extent to which a particular stimulus is a plausible cause of a given response. This suggests that though people may not be able to observe directly their cognitive processes, they will sometimes be able to report accurately about them. Accurate reports will occur when influential stimuli are salient and are plausible causes of the responses they produce, and will not occur when stimuli are not salient or are not plausible causes.

In short, people guess, and sometimes they get lucky. But where's the evidence?

Nisbett & Schachter, 1966. People were asked to get electric shocks to see how much shock they could stand (I myself would have waited to see if one of those see-how-much-free-

candy-you'll-eat studies from the post last week was still open). Half the subjects were also given a placebo pill which they were told would cause heart palpitations, tremors, and breathing irregularities - the main problems people report when they get shocked. The hypothesis: people who took the pill would attribute much of the unpleasantness of the shock to the pill instead, and so tolerate more shock. This occurred right on schedule: people who took the pill tolerated four times as strong a shock as controls. When asked why they did so well, the twelve subjects in the experimental group came up with fabricated reasons; one example given was "I played with radios as a child, so I'm used to electricity." Only three of twelve subjects made a connection between the pill and their shock tolerance; when the researchers revealed the deception and their hypothesis, most subjects said it was an interesting idea and probably explained the other subjects, but it hadn't affected them personally.

Zimbardo et al, 1965. Participants in this experiment were probably pleased to learn there were no electric shocks involved, right up until the point where the researchers told them they had to eat bugs. In one condition, a friendly and polite researcher made the request; in another, a surly and arrogant researcher asked. Everyone ate the bug (experimenters can be pretty convincing), but only the group accosted by the unpleasant researcher claimed to have liked it. This confirmed the team's hypothesis: the nice-researcher group would know why they ate the bug - to please their new best friend - but the mean-researcher group would either have to admit it was because they're pushovers, or explain it by saying they liked eating bugs. When asked after the experiment why they were so willing to eat the bug, they said things like "Oh, it's just one bug, it's no big deal." When

presented with the idea of cognitive dissonance, they once again agreed it was an interesting idea that probably affected some of the other subjects but of course not them.

Maier, 1931. Subjects were placed in a room with several interesting tools and asked to come up with as many solutions as possible to a puzzle about tying two cords together. One end of each cord was tied to the ceiling, and when the subject was holding on to one cord they couldn't reach the other. A few solutions were obvious, such as tying an extension cord to each, but the experiment involved a more complicated solution - tying a weight to a cord and using it as a pendulum to bring it into reach of the other. Subjects were generally unable to come up with this idea on their own in any reasonable amount of time, but when the experimenter, supposedly in the process of observing the subject, "accidentally" brushed up against one cord and set it swinging, most subjects were able to develop the solution within 45 seconds. However, when the experimenter asked immediately afterwards how they came up with the pendulum idea, the subjects were completely unable to recognize the experimenter's movement as the cue, and instead came up with completely unrelated ideas and invented thought processes, some rather complicated. After what the study calls "persistent probing", less than a third of the subjects mentioned the role of the experimenter.

Latane & Darley, 1970. This is the famous "bystander effect", where people are less likely to help when there are others present. The researchers asked subjects in bystander effect studies what factors influenced their decision not to help; the subjects gave many, but didn't mention the presence of other people.

Nisbett & Wilson, 1977. Subjects were primed with lists of words all relating to an unlisted word (eg "ocean" and "moon"

to elicit “tide”), and then asked the name of a question, one possible answer to which involved the unlisted word (eg “What’s your favorite detergent?” “Tide!”). The experimenters confirmed that many more people who had been primed with the lists gave the unlisted answer than control subjects (eg more people who had memorized “ocean” and “moon” gave Tide as their favorite detergent). Then they asked subjects why they had chosen their answer, and the subjects generally gave totally unrelated responses (eg “I love the color of the Tide box” or “My mother uses Tide”). When the experiment was explained to subjects, only a third admitted that the words might have affected their answer; the rest kept insisting that Tide was really their favorite. Then they repeated the process with several other words and questions, continuing to ask if the word lists influenced answer choice. The subjects’ answers were effectively random - sometimes they believed the words didn’t affect them when statistically they probably did, other times they believed the words did affect them when statistically they probably didn’t.

Nisbett & Wilson, 1977. Subjects in a department store were asked to evaluate different articles of clothing in a line. As usually happens in this sort of task, people disproportionately chose the rightmost object (four times as often as the leftmost), no matter which object was on the right; this is technically referred to as a “position effect”. The customers were asked to justify their choices and were happy to do so based on different qualities of the fabric et cetera; none said their choice had anything to do with position, and the experimenters dryly mention that when they asked the subjects if this was a possibility, “virtually all subjects denied it, usually with a worried glance at the interviewer suggesting they felt that they…were dealing with a madman”.

Nisbett & Wilson, 1977. Subjects watched a video of a teacher with a foreign accent. In one group, the video showed the teacher acting kindly toward his students; in the other, it showed the teacher being strict and unfair. Subjects were asked to rate how much they liked the teacher, and also how much they liked his appearance and accent, which were the same across both groups. Because of the halo effect, students who saw the teacher acting nice thought he was attractive with a charming accent; people who saw the teacher acting mean thought he was ugly with a harsh accent. Then subjects were asked whether how much they liked the teacher had affected how much they liked the appearance and accent. They generally denied any halo effect, and in fact often insisted that part of the reason they hated the teacher so much was his awful clothes and annoying accent - the same clothes and accent which the nice-teacher group said were part of the reason they *liked* him so much!

There are about twice as many studies listed in the review article itself, but the trend is probably getting pretty clear. In some studies, like the bug-eating experiment, people perform behaviors and, when asked why they performed the behavior, guess wrong. Their true reasons for the behavior are unclear to them. In others, like the clothes position study, people make a choice, and when asked what preferences caused the choice, guess wrong. Again, their true reasons are unclear to them.

Nisbett and Wilson add that when they ask people to predict how they would react to the situations in their experiments, people “make predictions that in every case were similar to the erroneous reports given by the actual subjects.” In the bystander effect experiment, outsiders predict the presence or absence of others wouldn’t affect their ability to help, and

subjects claim (wrongly) that the presence or absence of others didn't affect their ability to help.

In fact, it goes further than this. In the word-priming study (remember? The one with Tide detergent?) Nisbett and Wilson asked outsiders to predict which sets of words would change answers to which questions (would hearing "ocean" and "moon" make you pick Tide as your favorite detergent? Would hearing "Thanksgiving" make you pick Turkey as a vacation destination?). The outsiders' guesses correlated not at all with which words genuinely changed answers, but very much with which words the subjects guessed had changed their answers. Perhaps the subjects' answers looked a lot like the outsiders' answers because both were engaged in the same process: guessing blindly.

These studies suggest that people do not have introspective awareness to the processes that generate their behavior. They guess their preferences, justifications, and beliefs by inferring the most plausible rationale for their observed behavior, but are unable to make these guesses qualitatively better than outside observers. This supports the view presented in the last few posts: that mental processes are the results of opaque preferences, and that our own "introspected" goals and preferences are a product of the same machinery that infers goals and preferences in others in order to predict their behavior.

Secrets of the Eliminati

Anyone who does not believe mental states are ontologically fundamental - ie anyone who denies the reality of something like a soul - has two choices about where to go next. They can try reducing mental states to smaller components, or they can stop talking about them entirely.

In a utility-maximizing AI, mental states can be reduced to smaller components. The AI will have goals, and those goals, upon closer examination, will be lines in a computer program.

But in the blue-minimizing robot, its "goal" isn't even a line in its program. There's nothing that looks remotely like a goal in its programming, and goals appear only when you make rough generalizations from its behavior in limited cases.

Philosophers are still very much arguing about whether this applies to humans; the two schools call themselves reductionists and eliminativists (with a third school of wishy-washy half-and-half people calling themselves revisionists). Reductionists want to reduce things like goals and preferences to the appropriate neurons in the brain; eliminativists want to prove that humans, like the blue-minimizing robot, don't have anything of the sort until you start looking at high level abstractions.

I took a similar tack asking ksvanhorn's question in yesterday's post - how can you get a more accurate picture of what your true preferences are? I said:

> I don't think there are *true* preferences. In one situation you have one tendency, in another situation you have another tendency, and "preference" is what it looks like when you try to categorize tendencies. But categorization

> is a passive and not an active process: if every day of the week I eat dinner at 6, I can generalize to say "I prefer to eat dinner at 6", but it would be non-explanatory to say that a preference toward dinner at 6 caused my behavior on each day. I think the best way to salvage preferences is to consider them as tendencies currently in reflective equilibrium.

A more practical example: when people discuss cryonics or anti-aging, the following argument usually comes up in one form or another: if you were in a burning building, you would try pretty hard to get out. Therefore, you must strongly dislike death and want to avoid it. But if you strongly dislike death and want to avoid it, you must be lying when you say you accept death as a natural part of life and think it's crass and selfish to try to cheat the Reaper. And therefore your reluctance to sign up for cryonics violates your own revealed preferences! You must just be trying to signal conformity or something.

The problem is that not signing up for cryonics is also a "revealed preference". "You wouldn't sign up for cryonics, which means you don't really fear death so much, so why bother running from a burning building?" is an equally good argument, although no one except maybe Marcus Aurelius would take it seriously.

Both these arguments assume that somewhere, deep down, there's a utility function with a single term for "death" in it, and all decisions just call upon this particular level of death or anti-death preference.

More explanatory of the way people actually behave is that there's no unified preference for or against death, but rather a set of behaviors. Being in a burning building activates fleeing

behavior; contemplating death from old age does not activate cryonics-buying behavior. People guess at their opinions about death by analyzing these behaviors, usually with a bit of signalling thrown in. If they desire consistency - and most people do - maybe they'll change some of their other behaviors to conform to their hypothesized opinion.

One more example. I've previously brought up the case of a rationalist who knows there's no such thing as ghosts, but is still uncomfortable in a haunted house. So does he believe in ghosts or not? If you insist on there being a variable somewhere in his head marked $belief_in_ghosts = (0,1) then it's going to be pretty mysterious when that variable looks like zero when he's talking to the Skeptics Association, and one when he's running away from a creaky staircase at midnight.

But it's not at all mysterious that the thought "I don't believe in ghosts" gets reinforced because it makes him feel intelligent and modern, and staying around a creaky staircase at midnight gets punished because it makes him afraid.

Behaviorism was one of the first and most successful eliminationist theories. I've so far ignored the most modern and exciting eliminationist theory, connectionism, because it involves a lot of math and is very hard to process on an intuitive level. In the next post, I want to try to explain the very basics of connectionism, why it's so exciting, and why it helps justify discussion of behaviorist principles.

Tendencies in Reflective Equilibrium

Consider a case, not too different from what has been shown to happen in reality, where we ask Bob what sounds like a fair punishment for a homeless man who steals $1,000, and he answers ten years. Suppose we wait until Bob has forgotten that we ever asked the first question, and then ask him what sounds like a fair punishment for a hedge fund manager who steals $1,000,000, and he says five years. Maybe we even wait until he forgets the whole affair, and then ask him the same questions again with the same answers, confirming that these are stable preferences.

If we now confront Bob with both numbers together, informing him that he supported a ten year sentence for stealing $1,000 and a five year sentence for stealing $1,000,000, a couple of things might happen. He could say "Yeah, I genuinely believe poor people deserve greater penalties than rich people." But more likely he says "Oh, I guess I was prejudiced." Then if we ask him the same question again, he comes up with two numbers that follow the expected mathematical relationship and punish the greater theft with more jail time.

Bob isn't working off of some predefined algorithm for determining punishment, like "jail time = (10 * amount stolen)/net worth". I don't know if anyone knows exactly what Bob is doing, but at a stab, he's seeing how many unpleasant feelings get generated by imagining the crime, then proposing a jail sentence that activates about an equal amount of unpleasant feelings. If the thought of a homeless man makes images of crime more readily available and so increases the

unpleasant feelings, things won't go well for the homeless man. If you're really hungry, that probably won't help either.

So just like nothing automatically synchronizes the intention to study a foreign language and the behavior of studying it, so nothing automatically synchronizes thoughts about punishing the theft of $1000 and punishing the theft of $1000000.

Of course, there is something that non-automatically does it. After all, in order to elicit this strange behavior from Bob, we had to wait until he forgot about the first answer. Otherwise, he would have noticed and quickly adjusted his answers to make sense.

We probably could represent Bob's tendencies as an equation and call it a preference. Maybe it would be a long equation with terms for net worth of criminal, amount stolen, how much food Bob's eaten in the past six hours, and whether his local sports team won the pennant recently, with appropriate coefficients and powers for each. But if Bob saw this equation, he certainly wouldn't endorse it. He'd probably be horrified. It's also unstable: if given a choice, he would undergo brain surgery to remove this equation, thus preventing it from being satisfied. This is why I am reluctant to call these potential formalizations of these equations a "preference".

Instead of saying that Bob has one preference determining his jail time assignments, it would be better to model him as having several tendencies - a tendency to give a certain answer in the $1000 case, a tendency to give a different answer in the $1000000 case, and several tendencies towards things like consistency, fairness, compassion, et cetera.

People strongly consciously endorse these latter tendencies, probably because they're socially useful[1]. If the Chief of Police says "I know I just put this guy in jail for theft, but I'm

going to let this other thief off because he's my friend, and I don't really value consistency that much," then they're not going to stay Chief of Police for very long.

Bayesians and rationalists, in particular, make a big deal out of consistency. One common parable on the importance of consistency is the Dutch Book - a way to get free money from anyone behaving inconsistently. Suppose you have a weighted coin which can land on either heads or tails. There are several good reasons why I should not assign a probability of 66% to heads and 66% to tails, but one of the clearest is this: you can make me a bet that I will give you $2 if it lands on tails and you give me $1 if it lands on heads, and then a second bet where I give you $2 if it lands on heads and you give me $1 if it lands on tails. Whichever way the coin lands, I owe you $1 and you owe me $2 - I have gained a free dollar. So consistency is good if you don't want to be handing dollars out to random people…

…except that the Dutch book itself assumes consistency. If I believe that there is a 66% chance of it landing on heads, but refuse to take a bet at 2:1 odds - or even at 1.5:1 odds even though I should think it's easy money! - then I can't be Dutch booked. I am literally too stupid to be tricked effectively. You would think this wouldn't happen too often, since people would need to construct an accurate mental model to know when they should refuse such a bet, and such an accurate model would tell them they should revise their probabilities - but time after time people have demonstrated the ability to do exactly that.

I have not yet accepted that consistency is always the best course in every situation. For example, in Pascal's Mugging, a random person threatens to take away a zillion units of utility if you don't pay them $5. The probability they can make good

on their threat is miniscule, but by multiplying out by the size of the threat, it still ought to motivate you to give the money. Some belief has to give - the belief that multiplication works, the belief that I shouldn't pay the money, or the belief that I should be consistent all the time - and right now, consistency seems like the weakest link in the chain.

The best we can do is seek reflective equilibrium among our tendencies. If you endorse the belief that rich people should not get lighter sentences than poor people more strongly than you endorse the tendency to give the homeless man ten years in jail and the fund manager five, then you can edit the latter tendency and come up with a "fair" sentence. This is Eliezer's defense of reason and philosophy, a powerful justification for morality (see part one here) and it's probably the best we can do in justifying our motivations as well.

Any tendency that has reached reflective equilibrium in your current state is about as close to a preference as you're going to get. It still won't automatically motivate you, of course. But you can motivate yourself toward it obliquely, and come up with the course of action that you most thoroughly endorse.

FOOTNOTES:

1: A tendency toward consistency can cause trouble if someone gains advantage from both of two mutually inconsistent ideas. Trivers' hypothesis predicts that people will consciously deny the inconsistency so they can continue holding both ideas, yet still remain consistent and so socially acceptable. Rationalists are so annoying because we go around telling people they can't do that.

Hansonian Optimism

Imagine a kingdom ruled by a wise and benevolent king who, by reason of some strange tradition, is prohibited from ever leaving his palace. He only receives information on the affairs of the kingdom from his various Viziers. Like most Viziers, they are evil and power-hungry, and they are all conspiring with some of the most brutal and oppressive nobles in the land to preserve their reign of terror.

Also, you have an adorable pet bear

One day the Heroine speaks out against the current conditions in the kingdom. Taxes are too high, the peasants are starving to death, and people are being enslaved, all to enrich a few brutal nobles. The Heroine goes from town to town with her message: the people must beg the King to do something about this problem.

The Viziers hear of this and go to the king. “The Heroine,” they say, “is speaking against you. The whole kingdom is happy and prosperous, but this one woman wants to tear it apart and start a civil war for her own personal enrichment. Your people beg you to do something about her before she destroys the golden age they are currently enjoying.”

And so the king orders the Heroine executed, an order which the Viziers and the nobles are all too happy to carry out.

In this kingdom all the laws would be utterly selfish and show no regard for the average citizen. But this would be totally consistent with the King himself being a good person.

In fact, the King could be a *perfectly* good person, a person who attains moral heights of which other people never even dream, simply because he would never face a true moral dilemma. Suppose there were some problem that might prove morally difficult for the King – for example, the eastern states, which provide most of the kingdom's silk, are rebelling, and the king could either choose to live in peace with the newly independent east, or brutally crush them. If he chose the first option, silk would cost a lot more, and the king really likes silk.

If he were aware of this situation, it would be a sort of moral dilemma – do I do the right thing and avoid a war, or do I do what's convenient for me and lets me keep my luxury goods? Thanks to the Viziers, this problem disappears. If the Viziers want silk, they can tell the king that the eastern states have just launched a surprise attack, complete with atrocities – they must be dealt with as a matter of existential threat to the kingdom itself. And if the Viziers don't want silk, they can just tell the king that the east ran out of silk, too bad, nothing we can do about it. In fact, the Viziers will *never* present a true moral dilemma to the king, because then they wouldn't know which side he'd choose.

And so the king is faced only with easy, convenient moral decisions, and is able to preserve perfect innocence and purity. No matter how awful and tyrannical his decisions, the

populace may at least take consolation that their king is, at heart, a good person.

These were some of the thoughts that went through my head when I read Ozy's The Inherent Goodness of Human Nature, or Lack Thereof. Ozy worries that Hansonian explanations – in which people do nice things mostly for selfish reasons like signaling or self-signaling – mean that there's no such thing as goodness. As ze puts it:

> Robin Hanson (if I understand him correctly) would argue that the person giving money to the Make a Wish Foundation doesn't actually want to help sick children; they want to feel nice, like the sort of person who helps sick children, and– more importantly– they want everyone else to believe that they're nice people who help sick children.
>
> My initial reaction to this is "No! That's horrible! You terrible person!" Unfortunately, "you're a terrible person!" is not actually an argument that something is not true. My sense of revulsion at that idea is nothing more than a sign that I'm biased in favor of the "humans: basically nice" explanation.

Robin himself commented by saying:

> I love people, even if I don't think they are as good as they like to let on. I hope others can love me under the same conditions.

This seems like one of the wisest things I have ever heard, and restores just a little of my faith in humanity. But I think I'm more optimistic than Robin is.

Like Ozy, I believe human nature is basically good even though people's actions seem based on selfish and amoral motives. This is no more contradictory than the King being basically good, even though all his decrees will seem based on selfish and amoral motives. If the King has no access to accurate information, but can only make decisions based on information gleaned from biased sources, then the biases of those sources will be reflected in his words and deeds.

I cannot say why I identify other people with the Kings of their minds rather than with the Viziers of their minds (or with the creepy guys standing next to the king of their minds) save that this is who I feel I am in my mind, it is how I would like other people to see me, and so it seems both accurate and kind to see other people that way as well. Upon this view, people are good by nature, far better than their actions suggest, and it is really hard not to love and respect them.

This is not to say I think there's no such thing as evil. I would prefer that evil be something different than mere stupidity, something more than "Osama bin Laden was dumb enough to believe his mental Viziers when they told him becoming a terrorist mastermind was the right thing to do, poor guy", and indeed it seems there are lots of good stupid people and evil smart people, even lots of irrational good people and rational evil people. Although I have no clear answer, I think I would define evil as certain habits of mind which make it extremely easy for your Viziers to put one over on you, certain tendencies like "other people would do the same to me, so I'm just giving them a taste of my own medicine if I hurt them."

This is *still* an attempt to be a good person – it's an attempt to create a moral system in which you are just and virtuous for hurting others – but once you're letting your Viziers use this kind of argument on you I think it's pretty safe to say you've

gone evil. This doesn't quite correspond to my inner intuitive impression of evil, but if I turn it from a specific English-language assertion to a sort of preconscious sense of wrongedness and arrogant self-justification that expresses the same idea, it might.

You may notice how nicely this meshes with Trivers' theory of consciousness.

VIII. Doing Good

Newtonian Ethics

We often refer to morality as being a force; for example, some charity is “a force for good” or some argument “has great moral force”. But which force is it?

Consider the possibility that it is gravity. In statements like “Sentencing guidelines should take into account the gravity of the offense”, the words “gravity” and “immorality” are used interchangeably. Gravitational language informs our moral discourse in other ways too: immoral people are described as “fallen”, sin is a “weight” upon the soul, and we worry about society undergoing moral “collapse”. So the argument from common usage (is best argument! is never wrong!) makes a strong case for an unexpected identity between morality and gravity similar to that between (for example) electricity and magnetism.

We can confirm this to the case by investigating inverse square laws. If morality is indeed an unusual form of gravitation, it will vary with the square of the distance between two objects.

Imagine a village of a hundred people somewhere in the Congo. Ninety-nine of these people are malnourished, half-dead of poverty and starvation, oozing from a hundred infected sores easily attributable to the lack of soap and clean water. One of those people is well-off, living in a lovely two-story house with three cars, two laptops, and a wide-screen plasma TV. He refuses to give any money whatsoever to his ninety-nine neighbors, claiming that they’re not his problem. At a distance of a ten meters – the distance of his house to the nearest of their hovels – this is monstrous and abominable.

Now imagine that same hundredth person living in New York City, some ten thousand kilometers away. It is no longer monstrous and abominable that he does not help the ninety-nine villagers left in the Congo. Indeed, it is entirely normal; any New Yorker who spared too much thought for the Congo would be thought a bit strange, a bit with-their-head-in-the-clouds, maybe told to stop worrying about nameless Congolese and to start caring more about their friends and family.

This is, of course, completely rational. New York City, at ten thousand kilometers, is one million times further away from the suffering villagers as the original well-off man's ten meters. Since moral force decreases with the square of the distance, the moral force of the Congolese on the New Yorker is diminished by a factor of one million squared – that is, one trillion.

At that distance, all one billion Africans matter only 1/1000th as much as would a person at zero distance. There is, in fact, a person at zero distance from the average New Yorker – that New Yorker herself. So we find that our theory predicts that our obligations to the Congo are only one tenth of one percent as important as our obligations to ourselves.

We can confirm this experimentally. This article from 2005 lists private US overseas charitable contributions at $10.7 billion a year. The 2000 US Census gave a population of 281,421,906, meaning that the average American gave $38.02 in overseas charity. This is 0.107% of the average 2005 per capita income of $35,242, compared to a predicted .0100; that is, a margin of error of only about twenty four cents.

(This is why I love physics. You'd never get results that match up to predictions that precisely in the so-called "social

sciences".)

This methodology can be used to answer a seemingly very different problem that many of us face every day: just how far away from a beggar do you need to walk before you don't have to feel bad about not giving her money?

Suppose the marginal value of an extra dollar to a beggar is ten times its value to a well-off person such as yourself. We start with the money in your pocket, about a meter away from your brain. If you pass right by the beggar then the money may be a meter away from the beggar as well. Distance to both people is equal, so here the moral force exerted by the beggar is ten times stronger than your own moral force: you are clearly obligated to give her the money.

As you double your distance from the beggar to two meters, the moral force of her need decreases by a factor of four; however, she still has a 2.5x greater claim to the money than you do. Even three meters is not sufficient; her claim will be 1.1x as strong as your own.

However, four meters ought to do it. At this distance, the importance of the beggar's poverty has decreased by a factor of sixteen, while your own moral force has stayed constant. It's now 1.6x better for you to keep the money for yourself – a comfortable margin of safety.

There has been some discussion on whether it is acceptable to just hang to the far outside of the sidewalk in order to avoid a beggar, or whether this is unethical and it necessary to cross to the entire opposite side of the street. We now have the tools necessary to solve this problem. If you are on a commercial throughway, downtown residential, or other sidewalk listed on this table as having a minimum width of 4m or greater, it is borderline acceptable (ignoring air resistance) simply to move

to the other side of the walkway. However, on the smaller neighborhood residential sidewalks, industrial sidewalks and alleyways – not to mention anywhere the beggar is in the middle of the walkway – it is unfortunately necessary to cross all the way to the other side of the street.

Once again, the results of even a back-of-the-envelope calculation like this one mesh admirably with most people's native intuitions. Just as even a young child who throws a ball will have a "gut feeling" about how long it will stay up in the air, so even people unaware that morality is a variant of gravitation can correctly apply these same "gut feelings" to moral dilemmas.

In summary, morality is a form of gravitation, albeit an unusual one. Calculations performed based on inverse square law assumptions correctly predict most people's moral actions. Indeed, the majority of human moral behavior make no sense *except* under these assumptions, and without them our everyday moral reasoning would be ridiculous indeed.

Efficient Charity: Do Unto Others…

Imagine you are setting out on a dangerous expedition through the Arctic on a limited budget. The grizzled old prospector at the general store shakes his head sadly: you can't afford everything you need; you'll just have to purchase the bare essentials and hope you get lucky. But what is essential? Should you buy the warmest parka, if it means you can't afford a sleeping bag? Should you bring an extra week's food, just in case, even if it means going without a rifle? Or can you buy the rifle, leave the food, and hunt for your dinner?

And how about the field guide to Arctic flowers? You like flowers, and you'd hate to feel like you're failing to appreciate the harsh yet delicate environment around you. And a digital camera, of course - if you make it back alive, you'll have to put the Arctic expedition pics up on Facebook. And a hand-crafted scarf with authentic Inuit tribal patterns woven from organic fibres! Wicked!

…but of course buying any of those items would be insane. The problem is what economists call opportunity costs: buying one thing costs money that could be used to buy others. A hand-crafted designer scarf might have some value in the Arctic, but it would cost so much it would prevent you from buying much more important things. And when your life is on the line, things like impressing your friends and buying organic pale in comparison. You have one goal - staying alive - and your only problem is how to distribute your resources to keep your chances as high as possible. These sorts of economics concepts are natural enough when faced with a journey through the freezing tundra.

But they are decidedly not natural when facing a decision about charitable giving. Most donors say they want to "help people". If that's true, they should try to distribute their resources to help people as much as possible. Most people don't. In the "Buy A Brushstroke" campaign, eleven thousand British donors gave a total of £550,000 to keep the famous painting "Blue Rigi" in a UK museum. If they had given that £550,000 to buy better sanitation systems in African villages instead, the latest statistics suggest it would have saved the lives of about one thousand two hundred people from disease. Each individual $50 donation could have given a year of normal life back to a Third Worlder afflicted with a disabling condition like blindness or limb deformity..

Most of those 11,000 donors genuinely wanted to help people by preserving access to the original canvas of a beautiful painting. And most of those 11,000 donors, if you asked, would say that a thousand people's lives are more important than a beautiful painting, original or no. But these people didn't have the proper mental habits to realize that was the choice before them, and so a beautiful painting remains in a British museum and somewhere in the Third World a thousand people are dead.

If you are to "love your neighbor as yourself", then you should be as careful in maximizing the benefit to others when donating to charity as you would be in maximizing the benefit to yourself when choosing purchases for a polar trek. And if you wouldn't buy a pretty picture to hang on your sled in preference to a parka, you should consider not helping save a famous painting in preference to helping save a thousand lives.

Not all charitable choices are as simple as that one, but many charitable choices do have right answers. GiveWell.org, a site which collects and interprets data on the effectiveness of

charities, predicts that antimalarial drugs save one child from malaria per $5,000 worth of medicine, but insecticide-treated bed nets save one child from malaria per $500 worth of netting. If you want to save children, donating bed nets instead of antimalarial drugs is the objectively right answer, the same way buying a $500 TV instead of an identical TV that costs $5,000 is the right answer. And since saving a child from diarrheal disease costs $5,000, donating to an organization fighting malaria instead of an organization fighting diarrhea is the right answer, unless you are donating based on some criteria other than whether you're helping children or not.

Say all of the best Arctic explorers agree that the three most important things for surviving in the Arctic are good boots, a good coat, and good food. Perhaps they have run highly unethical studies in which they release thousands of people into the Arctic with different combination of gear, and consistently find that only the ones with good boots, coats, and food survive. Then there is only one best answer to the question "What gear do I buy if I want to survive" - good boots, good food, and a good coat. Your preferences are irrelevant; you may choose to go with alternate gear, but only if you don't mind dying.

And likewise, there is only one best charity: the one that helps the most people the greatest amount per dollar. This is vague, and it is up to you to decide whether a charity that raises forty children's marks by one letter grade for $100 helps people more or less than one that prevents one fatal case of tuberculosis per $100 or one that saves twenty acres of rainforest per $100. But you cannot abdicate the decision, or you risk ending up like the 11,000 people who accidentally decided that a pretty picture was worth more than a thousand people's lives.

Deciding which charity is the best is hard. It may be straightforward to say that one form of antimalarial therapy is more effective than another. But how do both compare to financing medical research that might or might not develop a "magic bullet" cure for malaria? Or financing development of a new kind of supercomputer that might speed up all medical research? There is no easy answer, but the question has to be asked.

What about just comparing charities on overhead costs, the one easy-to-find statistic that's universally applicable across all organizations? This solution is simple, elegant, and wrong. High overhead costs are only one possible failure mode for a charity. Consider again the Arctic explorer, trying to decide between a $200 parka and a $200 digital camera. Perhaps a parka only cost $100 to make and the manufacturer takes $100 profit, but the camera cost $200 to make and the manufacturer is selling it at cost. This speaks in favor of the moral qualities of the camera manufacturer, but given the choice the explorer should still buy the parka. The camera does something useless very efficiently, the parka does something vital inefficiently. A parka sold at cost would be best, but in its absence the explorer shouldn't hesitate to choose the the parka over the camera. The same applies to charity. An antimalarial net charity that saves one life per $500 with 50% overhead is better than an antidiarrheal drug charity that saves one life per $5000 with 0% overhead: $10,000 donated to the high-overhead charity will save ten lives; $10,000 to the lower-overhead will only save two. Here the right answer is to donate to the antimalarial charity while encouraging it to find ways to lower its overhead. In any case, examining the financial practices of a charity is helpful but not enough to answer the "which is the best charity?" question.

Just as there is only one best charity, there is only one best way to donate to that charity. Whether you volunteer versus donate money versus raise awareness is your own choice, but that choice has consequences. If a high-powered lawyer who makes $1,000 an hour chooses to take an hour off to help clean up litter on the beach, he's wasted the opportunity to work overtime that day, make $1,000, donate to a charity that will hire a hundred poor people for $10/hour to clean up litter, and end up with a hundred times more litter removed. If he went to the beach because he wanted the sunlight and the fresh air and the warm feeling of personally contributing to something, that's fine. If he actually wanted to help people by beautifying the beach, he's chosen an objectively wrong way to go about it. And if he wanted to help people, period, he's chosen a very wrong way to go about it, since that $1,000 could save two people from malaria. Unless the litter he removed is really worth more than two people's lives to him, he's erring even according to his own value system.

…and the same is true if his philanthropy leads him to work full-time at a nonprofit instead of going to law school to become a lawyer who makes $1,000 / hour in the first place. Unless it's one HELL of a nonprofit.

The Roman historian Sallust said of Cato "He preferred to be good, rather than to seem so". The lawyer who quits a high-powered law firm to work at a nonprofit organization certainly seems like a good person. But if we define "good" as helping people, then the lawyer who stays at his law firm but donates the profit to charity is taking Cato's path of maximizing how much good he does, rather than how good he looks.

And this dichotomy between being and seeming good applies not only to looking good to others, but to ourselves. When we donate to charity, one incentive is the warm glow of a job well

done. A lawyer who spends his day picking up litter will feel a sense of personal connection to his sacrifice and relive the memory of how nice he is every time he and his friends return to that beach. A lawyer who works overtime and donates the money online to starving orphans in Romania may never get that same warm glow. But concern with a warm glow is, at root, concern about seeming good rather than being good - albeit seeming good to yourself rather than to others. There's nothing wrong with donating to charity as a form of entertainment if it's what you want - giving money to the Art Fund may well be a quicker way to give yourself a warm feeling than seeing a romantic comedy at the cinema - but charity given by people who genuinely want to be good and not just to feel that way requires more forethought.

It is important to be rational about charity for the same reason it is important to be rational about Arctic exploration: it requires the same awareness of opportunity costs and the same hard-headed commitment to investigating efficient use of resources, and it may well be a matter of life and death. Consider going to www.GiveWell.org and making use of the excellent resources on effective charity they have available.

The Economics of Art and the Art of Economics

Here in Detroit, there is debate and concern over the possibility that the city's bankruptcy might obligate it to sell off masterpieces in the local art museum. Is solving a temporary financial problem really worth the cultural impoverishment of the city?

Yes. From Marginal Revolution:

> Consider "The Wedding Dance," a 16th-century work by the Flemish painter Pieter Bruegel the Elder. Detroit museum visitors have enjoyed this painting since 1930. How much would it cost to preserve that privilege for future generations?
>
> A tidy sum, as it turns out. According to Christie's, this canvas alone could fetch up to $200 million. Once interest rates return to normal levels — say, 6 percent — the forgone interest on that amount would be approximately $12 million a year.
>
> If we assume that the museum would be open 2,000 hours a year, and ignore the cost of gallery space and other indirect expenses, the cost of keeping the painting on display would be more than $6,000 an hour. Assuming that an average of five people would view it per hour, all year long, it would still cost more than $1,200 an hour to provide the experience for each visitor.

So the question of "should Detroit keep this painting?" reduces to "does the average visitor to the art museum derive $1200 in value from seeing this particular painting?" which is very

close to “would you pay $1200 for a ticket to an art museum that only had this painting in it?”

(other people may be more cultured than I am, but I find when I’m in an art museum I spend about ten seconds looking at each painting before moving on to the next one. So for me, at least, the cost is $120 per second of viewing time)

In If It’s Worth Doing, It’s Worth Doing With Made Up Statistics, I endorse trying to think quantitatively – not because we are always very good at quantifying things, but because sometimes just the attempt to quantify things makes the right answer so drop-dead obvious that whatever errors you make won’t change things one way or the other.

In the comments on MR people object that maybe some of the numbers in the calculation are a bit off, and that’s probably true. But just by trying the first numbers we think of, we realize we’re three orders of magnitude away from the spot where this would be a hard problem. And our numbers aren’t *that* off.

And this is why I continue to identify as consequentialist even though consequentialism is very hard and we can never do it exactly right. You don’t need a complete theory of ballistics in order to avoid shooting yourself in the foot.

Since I’m already being all soulless and analytical, let me just come out and say it – sell every piece of art in Detroit, but hire skilled forgers to make exact copies of them for a couple of hundred dollars each. You’ll have made billions of dollars, and the Detroit Art Museum will look exactly the same to anyone who’s not examining it through an electron microscope.

Sure, it’ll make it a little harder to signal snooty cultural superiority. But if you’re living in Detroit and trying to signal snooty cultural superiority, man, I don’t know what to tell you.

A Modest Proposal

I think dead children should be used as a unit of currency. I know this sounds controversial, but hear me out.

According to Population Services International, a respected charity research group, it costs between $650 and $1000 to save one person's life through charity. You've probably heard lower numbers like twenty cents somewhere. The lower numbers are wrong. Yes, maybe an anti-measles vaccine for a kid in Africa only costs twenty cents, and measles can be fatal. But there's a *lot* of overhead, and you have to immunize a *lot* of people before you get the one kid otherwise destined to die of measles. I find the $650-$1000 figure much more believable. Let's round it off to $800.

So one dead child = eight hundred dollars. If you spend eight hundred dollars on a laptop, that's one African kid who died because you didn't give it to charity. Distasteful but true. Now that we know that, we can get down to the details of designing the currency itself. It should be a big gold coin, with a picture of a smiling Burmese child on the front, and a tombstone on the back. The abbreviation can be DC.

Of course, most things won't cost a whole dead child, so we'll need smaller denominations. There are four dead puppies to a dead child, since dogs cost a bit above $200 to keep alive in an animal shelter. There are two acres of clear-cut rainforest per puppy., and five wounded Kenyans per clear-cut rainforest. I'm sure we can find talented artists to design the coins for all of these.

Yes, you grudgingly admit, such a system is technically feasible, but why in blue blazes would we want to replace our

reassuring green dollar bills graced with dignified ex-presidents, with *that*?

I leave that question to an article I read on the BBC site today: woman spends £250,000 on a luxury doghouse for her Great Danes complete with spa and plasma TV.

This *does* sound sort of ridiculous, but clearly it is not ridiculous enough. After all, at least one person thought it would be a good idea. Clearly, saying "doghouse that costs 250,000 pounds" does not carry the appropriate punch of "do not buy this."

And that's why I recommend switching to a dead-child-based currency. "Doghouse that costs 250,000 pounds" might not carry the proper punch. "Doghouse that costs 500 dead children" does. Using dead children as a unit of currency carries a built-in awareness of opportunity costs. Yes, you can buy that doghouse, if you *really* think it's more important than spending that same money to save five hundred Haitian kids' lives. Go on! Dogs watching plasma TV! That sounds *adorable*!

After reading an article about Mormon tithing practices, I am hopeful that the switch from dollars to DCs will destroy organized religion as well. It sounds plausible for a church to say it needs two million dollars to move to a larger building. It even sounds plausible when a pastor gets up there in front of his congregation and says that God really wants every family to just give whatever little bit they're able, so that they can all buy a better house of worship and praise God in a more fitting sanctuary. My old synagogue did this for years, and no one found anything wrong with it; my parents even donated quite a big chunk of money. If my rabbi'd had to say "We need

twenty-five hundred dead children to move to a sweeter pad", the gig would have been up.

Not like I am any saint myself. The past two years, I've spent about two dead puppies on books from Amazon.com alone. I am probably going to spend very close to a whole dead child to fly home for my two week winter break, and I spent ten dead children on my trip around the world this summer. I spent four wounded Kenyans on *fantasy map-making software*. But at least in the back of my mind I realize I'm doing it. Can the people who spend one dead kid and one dead puppy on the world's most expensive sundae say the same? What about the Japanese guy spending 1050 dead kids on a mobile phone strap?

One of America's top pro-life groups, Focus on the Family, spends two hundred thousand dead children a year pushing its message of conservatism and opposition to abortion. Take a second to fully appreciate the irony there.

I'm not saying these people don't have a right to spend their presumably hard-earned money on whatever they want. Of course they have that right. I am just saying that if we took the simple common sense step of changing our monetary denomination from dollars to dead children, maybe they'd want something different.

C'mon, I bet you a wounded Kenyan it'd work great.

The Life Issue

Unequally Yoked has an article about drone warfare, beginning with "I don't want to talk consequentialist tactics here".

So of course I immediately thought: "I wonder what the consequentialist tactics of drone warfare are".

According to the US government, between 2000 and 3000 people have been killed in the 8-year history of the drone warfare program. A group of independent journalists came up with between 2500 and 3300, so the numbers seem roughly correct. (source: Wikipedia)

It's very controversial what percent of these were real terrorists and what percent were civilians. The government claims an excellent track record of only hitting terrorists, but as Unequally Yoked points out, the government's definition of "terrorist" includes any male of military age in a conflict zone who can't be proven to be a non-terrorist.

Anti-drone organizations claim extremely high civilian casualty rates: for example, two Pakistani groups both claim that *most* of the 2000 deaths are civilians, Pakistani politicians routinely make claims that "100% of drone related deaths are civilians", and a Stanford study was cited as saying that ninety eight percent of casualties are innocent civilians (!)

However, this seems to be a misreading of the study, which actually says that only 2% of targets are *high level* militants. The study actually claims that between 474 and 881 deaths were civilians. If we take the average there of 700ish, and the average total of 2500ish, then we get a 72% terrorist to 28% civilian rate.

Other studies give similar numbers. The Bureau of Investigative Journalism guesses “at least 385 civilians”, which given our 2500 total speculation means <85% terrorist to >15% civilian. The New America Foundation says 80% terrorists, although their methodology seems to be going off newspaper reports which in turn probably go off the government which in turn goes off the questionable definition mentioned above. On the other hand, the *Long War Journal* specifies that they go off *Pakistani* media reports, and they get 94% terrorists.

The most plausible study I’ve seen comes from a very very brave group of Associated Press reporters who actually went into drone country and interviewed villagers after drone strikes. They estimated that about 70% of the drone casualties they investigated were terrorists, and that the number rises to about 90% if you discount a single disaster in which 40 civilians died.

So the responsible organizations seem to be converging on the 70-90% terrorist range. They also all seem to agree that the drones have been getting better in recent years and that a majority of casualties were in the early years of the program. Let’s take the middle of that range and say 80% terrorist, 20% civilian.

This is significantly fewer civilian deaths than conventional warfare. World War II had 33% soldiers to 66% civilians. Vietnam was probably about the same. The coalition side in the Iraq war got 66% soldiers to 33% civilians. The Israel invasion of Gaza (according to Israel) was 75% soldiers, 25% civilians, or (according to peace activist groups) 55% soliders, 45% civilians. So drone warfare’s reputation for being “surgical” is an overstatement but it is at least better than the usual invade-and-shoot methods. (source: Wikipedia)

80% terrorist to 20% civilian means the 2500 casualties include 2000 terrorists and 500 civilians. In order to talk about how bad this is, we need to decide whether we care if terrorists die or not. I don't have a good answer, so let's calculate this three ways.

The drone program has been going eight years, but only five of those have been very active.

If we don't care about terrorists, there have been about 100 civilian deaths per year.

If we care about terrorists only 25% as much as civilians, there have been about 200 combined deaths per year.

If we care about terrorists exactly as much as civilians, there have been about 500 deaths per year.

Aside from deaths, there are various other problems - for example one study points out the psychological trauma incurred by people in the areas involved knowing a plane could fly out of the sky and kill you at any moment. I have no idea how to quantify that, but as we'll see later, this isn't as big a problem as it sounds.

So the costs of the drone warfare program are 100 to 500 deaths per year, plus some unquantifiables.

What are the benefits?

Well, one goal is to prevent terrorism. Currently, there are 3000 terrorism-related deaths in Afghanistan per year plus another 2000 in Pakistan.

Another goal is to prevent Afghanistan from sliding into civil war. According to Wikipedia…

…according to Wikipedia, typing in "Afghan Civil War" gets you to a page called "War in Afghanistan: 1978 to Present",

which is really depressing, and which doesn't even have the information I'm looking for. But according to a sketchy site with no citations, the period of Afghan civil war from 1988 to 2001 caused 400,000 deaths, working out to about 30,000 per year.

Another way to look at a potential civil war in Afghanistan is to compare it to the worst period of factional violence in Iraq, which took place in 2006 and had almost 20,000 deaths a year. By coincidence Iraq's population is very close to the same as Afghanistan's, so the number translates pretty well. This also order-of-magnitude matches the observed deaths from the civil war in Afghanistan, so let's average them and say a civil war implies 25,000 casualties per year.

That leaves the unquantifiables. But I expect that things like panic, trauma, etc, follow deaths. Drone warfare causes trauma to those left behind, but terrorism also causes trauma to those left behind, and civil war definitely causes trauma. I can't imagine how much trauma, but it should be at least sort of proportional to the number of deaths in each branch.

So taking our third number, where we value terrorists exactly as much as civilians, and throwing away unquantifiables and using deaths as the only metric:

Drone warfare decreases total deaths if it reduces terrorist attacks in Afghanistan and Pakistan by at least 10%.

Drone warfare decreases total deaths if it cuts the chance of Afghanistan descending into civil war by even 2%.

These seem like *potentially* low bars. On a very simplistic view, killing 400 terrorists a year including a disproportionate number of major terrorist leaders seems pretty likely to decrease terrorism 10% if not further. And although it's hard to calculate exactly how much drone attacks prevent civil war, it

doesn't seem too crazy to think killing a bunch of rebels and rebel leaders would decrease that chance by at least 2%.

(if we don't care about terrorists' lives, then drone warfare is justified if it decreases total terrorist attacks by 2% or risk of civil war by 0.4%)

So in a very simplistic life-for-life calculus, drone warfare seems extremely defensible.

However, there are still some strong arguments that could be made against it:

1. Anger over drone warfare turns enough non-terrorists into terrorists, or lazy terrorists into actively plotting terrorists, that it indirectly *increases* the number of terrorist attacks or the chance of a civil war.

2. "Status quo plus drone warfare" versus "Status quo minus drone warfare" is a false dichotomy. There is some other more radical solution. For example, just withdraw completely and hope for the best, or even don't hope for the best but assume the civil war that will happen would have been inevitable anyway.

I don't think I like argument 2, although I don't know enough about it to really have a strong opinion. It seems like given how bad a civil war would be, any reasonable chance of averting it is sufficient reason to stay around. Argument 1 is much more potentially convincing, but I don't know how much so.

If I were the president, I would set up prediction markets on likelihood of civil war conditional on drone strikes, no drone strikes, and immediate retreat. I might also set one up on terrorist attacks per year conditional on each of those cases. Then I think I would actually have the information needed to

make an almost okay decision. Without that, I'm still agnostic about drone warfare. The most I can say is that I don't think one would have an easy time opposing it solely based on the direct death toll. This is actually not the conclusion I was expecting, so please check my calculations and see if I did anything wrong.

Now I don't think Leah *thinks* I'm contradicting her article, because she says she's potentially sympathetic to consequentialist calculations. All she wants is to realize the enormity of their decision before acting:

> *I don't want to talk consequentialist tactics here or ticking timebomb scenarios. Whether or not you support sanctions, we have a duty to talk about them without euphemisms. Our politicians should face up to the enormity of the violence they plan to inflict on others, not puff themselves up by telling us how strong they are, how able and happy they are to make other people destitute or dead.*
>
> *…if we don't label these actions as warped and unnatural when we perform them out of necessity, we might forget the enormity of our transgression. And, if we do faithfully name them as they are, we might find that fewer of them seem all that necessary.*

And I am sympathetic to this. We should *always* consider the importance and human cost of moral decisions. We should always wish that we could save everyone - although I don't know if actual guilt about it is very healthy.

But what I reject is the implicit idea that this should be one-sided. The president who decides to launch a drone attack in

order to save people from terrorism later on should have to think long and hard about what he's doing - to really imagine the civilians who might die, and the pain of their families, instead of thinking of them as "collateral damage".

But the equal and opposite president who decides *not* to launch the drone attack should have to think long and hard about what he's doing too - to really take on board that if that terrorist he decided not to kill blows himself up in a busy marketplace two weeks later killing forty people, all those deaths are now on his conscience.

I think Obama has gone through enough hand-wringing that I'm prepared to give him a pass on this one. I hope his critics understand they need some hand-wringing too.

What if Drone Warfare Had Come First?

Epistemic Status: *Interesting to think about, but not nearly as aimed at expressing a strong position on this issue as it might sound.*

I am somewhat happy that no one has torn my calculations apart on the drone warfare article yet. Only somewhat happy because I hoped someone would try and I would get either independent confirmation or competing data to take into account.

But several people did respond, and the overall tone was that drone warfare has more problems than just raw death count. It's dehumanizing. It makes warfare "too easy" and hides the real cost. It gives too much power to whoever makes drone-related decisions. It violates the rules of war.

These are all good points. But I can't help but think back to the old Less Wrong article If Many Worlds Had Come First. It's sort of about quantum mechanics, but it's also about the dangers of applying higher standards to later innovations than to entrenched conventional wisdom.

There are sometimes strong arguments for doing this. For example, doctors often prescribe older, apparently-worse drugs over newer, apparently-better drugs (especially in pregnancy) just because they feel like they already know the side effects of the older drugs whereas the newer drugs might have side effects that are yet to reveal themselves. This model certainly has implications for drone warfare: it looks good now, but we don't know the long-term effects.

Still, in the spirit of that Less Wrong article, I can't help but wonder what people would think if drone warfare had come

first:

The scene is the Oval Office. Three of the Joint Chiefs of Staff, GENERAL HAWKE, GENERAL STEELE, and GENERAL RIPPER, are meeting with THE PRESIDENT. The meeting has been a long and exhausting discussion of drone strikes, and they are reaching the end.

PRESIDENT: I think we only have one more matter left to discuss. As you know, I have recently been worried about the moral cost of our drone war. So many lives lost. So many civilian casualties. I tasked DARPA with coming up with a *new* type of warfare, one which will end some of the troubling moral quandaries with which we are forced to wrestle every day. I believe General Ripper has been briefed on the results?

HAWKE: Mr. President, once again, I object to this pie-in-the-sky project. Drone warfare was good enough for our ancestors and it is good enough for us. The Romans used surgically precise ballista strikes to assassinate Hannibal without harming the Carthaginian populace. Abraham Lincoln used guided hot-air balloons to knock out top Confederate officials and keep this country united. Literally *hundreds* of people died in World War I before the British were finally able to kill Kaiser Wilhelm with a carefully-aimed zeppelin. To abandon drone warfare now for some untested new project would be an insult to their memory!

PRESIDENT: General Hawke, I appreciate your concerns, and I promise I will not be overly hasty to embrace these new ideas. But I'd like to hear what General Ripper has to say.

RIPPER: *(interjecting)* Guys!…Guys! Guys, listen! This is going to be *so awesome*. Listen to this! We take hundreds of thousands of people…guys, listen!…we take hundreds of thousands of people, give them really really really powerful

automatic weapons…this is going to be so awesome…we take hundreds of thousands of people and give them really powerful automatic weapons and put them on planes and give them parachutes and drop them into our enemies' cities and then they just start shooting everything BLAM BLAM BLAM until our enemies run away and we're like HA HA HA HA HA THIS IS OUR CITY NOW and then we win!

STEELE: What the hell, Ripper?

RIPPER: No, listen, this will totally work! We take hundreds of thousands of people. We can use young kids and poor people and minorities, because we don't have to pay them as much. And then we give them really really big weapons. Like, not just the kinds of guns hunters use. Not even the kind of guns we give police. Guns that just NEVER STOP SHOOTING BULLETS! You can just swing them in a big arc and it will leave an arc of bullets everywhere and *anyone anywhere in that arc will be dead*! It will be SO AWESOME!

HAWKE: Ripper, are you *mad?*

RIPPER: Guys, think about it! You're Ayatollah Sistani, or Mullah Omar, or one of those motherf@*kers. You're having breakfast in your house one day when WHAM! A hundred thousand American teenagers and minorities RIGHT IN YOUR CITY with guns that never stop shooting bullets! There are bullet holes in your walls and in your gardens and now they're shooting your water supply and your power plant and *everything*. Do you think you're going to keep having your f@*king breakfast? Or do you think you're going to start waving an American flag and get on board with American policies like, *right* away?

PRESIDENT: General Ripper, frankly your idea seems *at best* ill-advised! Just to take one of many objections, we'll

never be able to gather a hundred thousand Americans in secret. Ayatollah Sistani will hear about our plan long before we can surprise him.

RIPPER: And what could that motherf@*ker do about it?

STEELE: Well, he could get some Iranian teenagers and minorities, give *them* these super-guns of yours, and have them lie in wait for *our* teenagers and minorities outside his house.

RIPPER: Oh my god that would be so awesome! Because we have more technology, so we could have better guns than they do! And we're richer than they are, so we could hire more teenagers and minorities! Right? RIGHT? So everyone would be like BLAM BLAM BLAM with their super-guns and there would be this huge fight and in the end we would win and get that sunavab*tch anyway!

PRESIDENT: *(horrified)* You realize what you're suggesting is the deaths of dozens of Americans and Iranians, right? Maybe even hundreds!

RIPPER: No, look. It would be okay. Listen to this. We would come up…we would come up with this new philosophy where once a teenager or minority got a super-powerful gun from our enemies, it would be *okay* if we killed them. Because if we didn't kill them, they might use that gun to shoot us.

HAWKE: But they're only doing that because otherwise we would…I can't believe I have to say this…otherwise we would parachute teenagers with giant guns into their city to shoot the ayatollah.

RIPPER: I KNOW RIGHT? We're going to parachute teenagers with giant guns into their city to shoot the ayatollah!

THEN EVERYTHING'S GOING TO GET BLOWN UP AND IT'S GOING TO BE SO COOL.

STEELE Everything…blown up?

RIPPER: Oh man I totally forgot this part! If we just have the super guns, people might hide inside buildings, right? And then we couldn't shoot them and then the ayatollah wouldn't have to agree to do everything we say. So…ohmigod you guys are going to love this…we take cars, right? And we cover them in armor and put giant caterpillar tracks on the bottom so they can drive over walls and sh*t. And then we put HUMONGOUS GUNS on top of the cars. Guns so big they can BLOW UP WHOLE BUILDINGS. And then we just KEEP BLOWING UP THE CITY until the Ayatollah agrees to do everything we want.

PRESIDENT: *(to buzzer under desk, in a whisper)* Uh, Secret Service? One of the Joint Chiefs of Staff has started acting *really weird*. Maybe you could stand outside the door and, uh, monitor the situation?

RIPPER: And then! And then we have these planes, right? And we arm them with lots of bombs, and we fly them over enemy cities, and…

HAWKE: Oh, thank goodness. You're starting to see sense and admit that the old ways of drone warfare are right after all.

RIPPER: No, it would be totally different! Because, get this! There would be *people* in these planes! We'd train them at special schools and whirl them around in centrifuge until they were able to work at 5 g-forces without passing out. Whirl! Whirl! Whirl! And sometimes they'd bomb our enemies, and sometimes our enemies would shoot them down and they'd get captured and we'd have to send in special teams of super-spies

to rescue them before they got tortured and told our enemies everything they know!

STEELE That's…horrible!

RIPPER: And instead of trying to only target high-profile enemy leaders? We'd have a special rule that they *couldn't* target high-profile enemy leaders! They would have to hit power plants and dams and weapons factories and…

PRESIDENT: Weapons factories? Wouldn't those explode if bombed?

RIPPER: OH yeah. HUGE explosion! BOOM! And then when everything had been destroyed from the air, we could send in our hundred thousand teenagers with super guns and they could send in *their* hundred thousand teenagers with super guns, and we could send in our cars covered in metal with caterpillar treads and they could send in *their* cars covered in metal in caterpillar treads and then it would be all BLAM BLAM BLAM for WEEKS AND WEEKS and we win would because we would both kill each other and destroy each other's cars but we're bigger so we would have more of them and the Ayatollah would have to agree to do everything we say.

STEELE What if he doesn't?

RIPPER: We could kick him out, and say okay, city, you're part of America now! You're following American laws! You fly the American flag! And then America would be even bigger! And we could take their stuff too, like if there was any oil in the city, then it would be our oil!

PRESIDENT: General Ripper, this is highly unorthodox but I am going to have to relieve you of command effective immediately. This so-called "plan" of DARPA and yourself

appears to be no more than the rantings of a deranged and homicidal lunatic. Your request to further develop this new type of warfare is completely denied, and honestly you seem to have so little regard for human life or the rules of warfare that I do not want you anywhere near our nation's drone fleet.

STEELE: Wait, I just realized something. Maybe this isn't about having little regard for human life. Maybe it could even *help preserve* human life?

PRESIDENT: *(skeptically)* What do you mean?

STEELE: Think about it. Nowadays, our drone controllers plan strikes from the safety of the Pentagon, never knowing the horrors of warfare, never seeing their victims as real people. But imagine what would happen if we did war Ripper's way?

HAWKE: What would happen?

STEELE: All our teenagers and minorities would see the looks on the faces of their victims as they got shot. Reporters would go into the cities and televise the devastation that our cars with armor and humongous guns had caused. People would come back traumatized, and we'd see them and understand their trauma and with it the trauma of warfare.

PRESIDENT: And?

STEELE: And we'd only need to do it once. Think of the hundreds of people who died in World War I, Mr. President. Think about the waste. If we had done things Ripper's way, the Allies would have *encountered* the Germans. They would have realized they were human beings just like them. The people in the capitals would have had to think twice about sending their young men off to die just because they wanted to play stupid games with the balance of power. And they *would*

have thought twice. They would have said “No, this is horrible”. Instead of those hundreds of zeppelin-related casualties, we would have had both sides pull back from the brink of war, and join together in their common humanity. It would have been a War to End Wars.

HAWKE: It would never have happened that way.

STEELE: No, perhaps not. Perhaps we should go on with our drone strikes as usual. Keep killing hundreds of people. But perhaps one day we will regret not taking hundreds of thousands of teenagers from disadvantaged backgrounds, arming them with guns, parachuting them into our enemies’ cities, and having them shoot things until our enemies agree to do whatever we say. Maybe it will end up being the only truly virtuous mode of warfare, the only one that preserves our inherent humanity.

PRESIDENT: *(to buzzer under desk, in a whisper)* Yes, I’m sorry, the Joint Chiefs of Staff seem to have gone insane. Would you mind terribly coming in and escorting them out?

The Secret Service comes in and escorts the Joint Chiefs of Staff out. The President sighs and starts taking care of some paperwork. A few minutes later, MS. WELLS, the Secretary of Health and Human Services, comes in.

WELLS: Mr. President? I’m sorry to disturb you, but a question has come up. I know you authorized free health care for everyone in the nation, but the doctors are wondering whether it’s okay if they buy examination tables made of solid gold. Something about it ‘adding a touch of class to the clinic’.

PRESIDENT: Sure. Tell them to go ahead. We have more tax money than we know what to do with these days anyway.

Nefarious Nefazodone and Flashy Rare Side-Effects

[Epistemic status: I am still in training. I am not an expert on drugs. This is poorly-informed speculation about drugs and it should not be taken seriously without further research. Nothing in this post is medical advice.]

I.

Which is worse – ruining ten million people's sex lives for one year, or making one hundred people's livers explode?

I admit I sometimes use this blog to speculate about silly moral dilemmas for no reason, but that's not what's happening here. This is a real question that I deal with on a daily basis.

SSRIs, the class which includes most currently used antidepressants, are very safe in the traditional sense of "unlikely to kill you". Suicidal people take massive overdoses of SSRIs all the time, and usually end up with little more than a stomachache for their troubles. On the other hand, there's increasing awareness of very common side effects which, while not disabling, can be pretty unpleasant. About 50% of users report decreased sexual abilities, sometimes to the point of total loss of libido or anorgasmia. And something like 25% of users experience "emotional blunting" and the loss of ability to feel feelings normally.

Nefazodone (brand name Serzone®, which would also be a good brand name for a BDSM nightclub) is an equally good (and maybe better) antidepressant that does not have these side effects. On the other hand, every year, one in every 300,000 people using nefazodone will go into "fulminant hepatic failure", which means their liver suddenly and spectacularly stops working and they need a liver transplant or else they die.

There are a lot of drug rating sites, but the biggest is Drugs.com. 467 Drugs.com users have given Celexa, a very typical SSRI, an average rating of 7.8/10. 14 users have given nefazodone an average rating of 9.1/10.

CrazyMeds might not be as dignified as Drugs.com, but they have a big and well-educated user base and they're psych-specific. Their numbers are 3.3/5 (n = 253) for Celexa and 4.1/5 (n = 47) for nefazodone.

So both sites' users seem to agree that nefazodone is notably better than Celexa, in terms of a combined measure of effectiveness and side effects.

But nefazodone is practically never used. It's actually illegal in most countries. In the United States, parent company Bristol-Myers Squibb (which differs from normal Bristol-Myers in that it was born without innate magical ability) withdrew it from the market, and the only way you can find it nowadays is to get it is from an Israeli company that grabbed the molecule after it went off-patent. In several years working in psychiatry, I have never seen a patient on nefazodone, although I'm sure they exist somewhere. I would estimate its prescription numbers are about 1% of Celexa's, if that.

The problem is the hepatic side effects. Nobody wants to have their liver explode.

But. There are something like thirty million people in the US on antidepressants. If we put them all on nefazodone, that's about a hundred cooked livers per year. If we put them all on SSRIs, at least ten million of them will get sexual side effects, plus some emotional blunting.

My life vastly improved when I learned there was a searchable database of QALYs for different conditions. It doesn't have SSRI-induced sexual dysfunction, but it does have sexual

dysfunction due to prostate cancer treatment, and I assume that sexual dysfunction is about equally bad regardless of what causes it. Their sexual dysfunction has some QALY weights averaging about 0.85. Hm.

Assume everyone with fulminant liver failure dies. That's not true; some get liver transplants, maybe some even get a miracle and recover. But assume everyone dies – and further, they die at age 30, cutting their lives short by fifty years.

In that case, putting all depressed people on nefazodone for a year costs 5,000 QALYs, but putting all depressed people on SSRIs for a year costs 1,500,000 QALYs. The liver failures may be flashier, but the 3^^^3 dust specks worth of poor sex lives add up to more disutility in the end.

I don't want to overemphasize this particular calculation for a couple of reasons. First, SSRIs and nefazodone both have other side effects besides the major ones I've focused on here. Second, I don't know if the level of SSRI-induced sexual dysfunction is as bad as the prostate-surgery-induced sexual dysfunction on the database. Third, there are a whole bunch of antidepressants that are neither SSRIs nor nefazodone and which might be safer than either.

But I *do* want to emphasize this pattern, because it recurs again and again.

II.

In that spirit, which would you rather have – something like a million people addicted to amphetamines, or something like ten people have their skin eat itself from the inside?

I can't get good numbers on how many adults abuse Adderall, but a quick glance at the roster for my hospital's rehab unit suggests "a lot". Huffington Post calls it the most abused

prescription drug in America, which sounds about right to me. Honestly there are worse things to be addicted to than Adderall, but it's not completely without side effects. The obvious ones are anxiety, irritability, occasionally frank psychosis, and sometimes heart problems – but a lot of the doctors I work with go beyond what the research can really prove and suggest it can produce lasting negative personality change and predispose people to other forms of addictive and impulsive behavior.

If you've got to give adults a stimulant, I would much prefer modafinil. It's not addictive, it lacks most of Adderall's side effects, and it works pretty well. I've known many people on modafinil and they give it pretty universally positive reviews.

On the other hand, modafinil *may or may not* cause a skin reaction called Stevens Johnson Syndrome/Toxic Epidermal Necrolysis, which like most things with both "toxic" and "necro" in the name is really really bad. The original data suggesting a connection came from kids, who get all sorts of weird drug effects that adults don't, but since then some people have *claimed* to have found a connection with adults. Some people get SJS anyway just by bad luck, or because they're taking other drugs, so it's really hard to attribute cases specifically to modafinil.

Gwern's Modafinil FAQ mentions an FDA publication which argues that the background rate of SJS/TEN is 1-2 per million people per year, but the modafinil rate is about 6 per million people per year. However, there are only three known cases of a person above age 18 on modafinil getting SJS/TEN, and this might not be different from background rates after all. Overall the evidence that modafinil increases the rate of SJS/TEN in adults at all is pretty thin, and if it does, it's as rare as hen's

teeth (in fact, very close to the same rate as liver failure from nefazodone).

(also: consider that like half of Silicon Valley is on modafinil, yet San Francisco Bay is not yet running red with blood.)

(also: ibuprofen is linked to SJS/TEN, with about the same odds ratio as modafinil, but nobody cares, and they are correct not to care.)

I said I've never seen a doctor prescribe nefazodone in real life; I can't say that about modafinil. I have seen one doctor prescribe modafinil. It happened like this: a doctor I was working with was very upset, because she had an elderly patient with very low energy for some reason, I can't remember, maybe a stroke, and wanted to give him Adderall, but he had a heart arrythmia and Adderall probably wouldn't be safe for him.

I asked "What about modafinil?"

She said, "Modafinil? Really? But doesn't that sometimes cause Stevens Johnson Syndrome?"

And then I glared at her until she gave in and prescribed it.

But this is very, very typical. Doctors who give out Adderall like candy have no associations with modafinil except "that thing that sometimes causes Stevens-Johnson Syndrome" and are afraid to give it to people.

III.

Nefazodone and modafinil are far from the only examples of this pattern. MAOIs are like this too. So is clozapine. If I knew more about things other than psychiatry, I bet I could think of examples from other fields of medicine.

And partially this is natural and understandable. Doctors swear an oath to "first do no harm", and toxic epidermal necrolysis is pretty much the epitome of harm. Thought experiments like torture vs dust specks suggest that most people's moral intuitions say that *no* amount of aggregated lesser harms like sexual side effects and amphetamine addictions can equal the importance of avoiding even a tiny chance of some great harm like liver failure or SJS/TEN. Maybe your doctor, if you asked her directly, would endorse a principled stance of "I am happy to give any number of people anxiety and irritability in order to avoid even the smallest chance of one case of toxic epidermal necrolysis."

And yet.

The same doctors who would never *dare* give nefazodone, consider Seroquel a perfectly acceptable second-line treatment for depression. Along with other atypical antipsychotics, Seroquel raises the risk of sudden cardiac death by about 50%. The normal risk of cardiac sudden death in young people is about 10 in 100,000 per year, so if my calculations are right, low-dose Seroquel causes an extra cardiac death once per every 20,000 patient-years. That's ten times as often as nefazodone causes an extra liver death.

Yet nefazodone was taken off of the market by its creators and consigned to the dustbin of pharmacological history, and Seroquel is the sixth-best-selling drug in the United States, commonly given for depression, simple anxiety, and sometimes even to help people sleep.

Why the disconnect? Here's a theory: sudden cardiac death happens all the time; sometimes God just has it in for you and your heart stops working and you die. Antipsychotics can increase the chances of that happening, but it's a purely

statistical increase, such that we can detect it aggregated over large groups but never be sure that it played a role in any particular case. The average person who dies of Seroquel never knows they died of Seroquel, but the average person who dies from nefazodone is easily identified as a nefazodone-related death. So nefazodone gets these big stories in the media about this young person who died by taking this exotic psychiatric drug, and it becomes a big deal and scares the heck out of everybody. When someone dies of Seroquel, it's just an "oh, so sad, I guess his time has come."

But the end result is this. When treatment with an SSRI fails, nefazodone and Seroquel naively seem to be equally good alternatives. Except nefazodone has a death rate of 1/300,000 patient years, and Seroquel 1/20,000 patient years. And yet everyone stays the hell away from the nefazodone because it's known to be unsafe, and chooses the Seroquel.

I conclude either doctors are terrible at thinking about risk, or else maybe a little *too* good at thinking about risk.

I bring up the latter option because there's a principal-agent problem going on here. Doctors want to do what's best for their patients. But they also want to do what's best for themselves, which means not getting sued. No one has ever sued their doctor because they got a sexual side effect from SSRIs, but if somebody dies because they're the lucky 1/300,000 who gets liver failure from nefazodone, you can bet their family's going to sue. Suddenly it's not a matter of comparing QALYs, it's a matter of comparing zero percent chance of lawsuit with non-zero percent chance of lawsuit.

(Fermi calculation: if a doctor has 100 patients at a time on antidepressants, and works for 30 years, then if she uses Serzone as her go-to antidepressant, she's risking a 1% chance

of getting the liver failure side effect once in her career. That's small, but since a single bad lawsuit can bankrupt a doctor, it's worth taking seriously.)

And that would be a tough lawsuit to fight. "Yes, Your Honor, I knew when I prescribed this drug that it sometimes makes people's livers explode, but the alternative often gives people a bad sex life, and according to the theory of utilitarianism as propounded by 18th century philosopher Jeremy Bentham – " … "Bailiff, club this man".

And the same facet of nefazodone that makes it exciting for the media makes it exciting for lawsuits. When someone dies of nefazodone toxicity, everyone knows. When someone dies of Seroquel, "oh, so sad, I guess his time has come".

That makes Seroquel a lot safer than nefazodone. Safer for the doctor, I mean. The *important* kind of safer.

This is why, as I mentioned before, I hate lawsuits as a de facto regulatory mechanism. Our de jure regulatory mechanism, the FDA, is pretty terrible, but to its credit it hasn't banned nefazodone. One time it banned clozapine because of a flashy rare side effect, but everyone yelled at them and they apologized and changed their mind. With lawsuits there's nobody to yell at, so we just end up with people very quietly adjusting their decisions in the shadows and nobody else being any the wiser.

I don't want to overemphasize this, because I think it's only one small part of the problem. After all, a lot of countries withdrew nefazodone entirely and didn't even give lawsuits a chance to enter the picture.

But whatever the cause, the end result is that drugs with rare but spectacular side effects get consistently underprescribed relative to drugs with common but merely annoying side

effects, or drugs that have more side effects but manage to hide them better.

The Consequentialism FAQ

PART ZERO: INTRODUCTION

0.1: Who are you? Where am I?

You can find more about me at www.raikoth.net. This is the Consequentialist FAQ.

0.2: So what's all this then?

Consequentialism is a moral theory, i.e. a description of what morality means and how to solve moral problems. Although there are several explanations of it online, they're all very philosophical, which means they love to define terms and debate details and finally conclude that it is an important issue which no doubt will need to be meticulously deconstructed for several more centuries. This FAQ is intended for a different purpose. It is meant to convince you that consequentialism is the *right* moral system, and that all other moral systems are subtly but distinctly insane.

I do not claim full credit for the insights expressed in here. Most come from a long tradition of moral philosophers, and some of the more clever insights and turns of phrase come from the Less Wrong Metaethics Sequence.

0.3: Why?

The basic thesis is that consequentialism is the only system which both satisfies our moral intuition that morality should make a difference to the real world, and that we should care about other people. Other moral systems are more concerned with looking good than being good, and although this is not immediately apparent it will hopefully become clearer on closer inspection.

0.4: And who cares?

Part Eight will get into this further, but the basic summary is: we live in a failed world. Problems like world hunger, war, racism, and environmental damage are only partly controlled even in our insulated First World countries, and in the majority of the world they are barely controlled at all. It is traditional to attribute this to "people being immoral", but in fact people are generally very moral: they feel intense moral outrage at the suffering in the world, they are extremely generous in response to certain obvious opportunities for generosity like the Haitian earthquake, and many people will, in an emergency that calls for it, sacrifice their lives to save others with only a split second's thought. And even things that are in fact repulsive, like the intensity with which people oppose gay marriage, derive from a misplaced sense that they are doing the right and moral thing; people will devote their entire careers to opposing gay marriage even though it does not hurt them personally because they feel like they *should*. The problem isn't that people aren't trying to be moral, it's that they're no good at it. This FAQ tries to explain how to do it better.

0.5: Is this FAQ exhaustive?

No. This only provides a very quick introduction to consequentialism and why you should believe it. There are many concepts necessary in order to do consequentialism *right* - including game theory, decision theory, and some philosophy of law - that are barely touched upon or not even mentioned. These may change the results of important moral questions. All this FAQ claims to be useful for is to help get some basic intuitions right; figuring out how to translate those intuitions into action requires more work.

0.6: What is the structure of this FAQ?

Part One talks about what it means to philosophize about morality and solve moral dilemmas, though it is not intended as a full substitute for a real meta-ethical theory, which would be much more boring and interminable. Part Two introduces and defends the intuition that morality should have something to do with the real world. Part Three introduces and defends the intuition that morality should care about other people. Part Four finally gets to consequentialism and Part Five gets to its most famous example, utilitarianism. Part Six gets into rules and human rights, Part Seven clears up some common objections and thought experiments, and Part Eight sets out why I think this is really important and might save the world.

PART ONE: WHIRLWIND METAETHICS

1.1: What does it mean to search for moral rules?

Searching for moral rules means searching for principles that correctly describe and justify enough of our existing moral intuition that we feel confident applying them to decide edge cases.

There are many moral situations where nearly everyone agrees on the correct answer, even though we're not exactly sure why. For example, even if we don't have a formal theory of morality we know that killing an innocent person for no reason is morally wrong.

There are other moral situations in which there is wide disagreement on the morally correct answer: for example, is it acceptable to use the legal apparatus of the state to prevent women from aborting their unborn babies?

When arguing about this latter question, people try to appeal to existing moral principles that are widely agreed upon. For example, a pro-lifer might argue that we all agree on the moral

intution that it is wrong to take a life, and abortion takes a life, and therefore abortion is wrong by agreed moral rules. But a pro-choicer might argue that we all agree on the moral intuition that people should have control of their own bodies, and control over whether to abort a fetus is related to control over one's own body, and therefore abortion is acceptable by agreed moral rules.

Judging by the continued popularity of the abortion debate, this method is insufficient to quickly resolve moral edge cases.

To search for moral rules means to come up with a more formalized method of translating moral intutions into moral rules and applying those rules to edge cases, one which is clearly correct and which cannot be countered by an equal and opposite method of applying moral rules to edge cases.

1.2: Why care about moral intuitions?

Moral intuitions are people's basic ideas about morality. Some of them are hard-coded into the design of the human brain. Others are learned at a young age. They manifest as beliefs ("Hurting another person is wrong"), emotions (such as feeling sad whenever I see an innocent person get hurt) and actions (such as trying to avoid hurting another person.)

Moral intutions are important because unless you are a very specific type of philosopher they are the only reason you believe morality exists at all. They are also the standards by which you judge all moral philosophies; if the only content of a certain moral philosophy was "it's wrong to wear green clothes on Saturday", then you would not find this moral philosophy attractive unless it could justify itself by saying why wearing green clothes on Saturday affected other things that our moral intutions find more important. For example, if every time someone wore green clothes on Saturday, the world

become a safer and happier place, then the suggestion to wear green clothes on Saturday might seem justified - but in this case the work is being done by a moral intuition in favor of a safer and happier world, not by anything about green clothes themselves. On the other hand, if a philosopher were to justify a moral theory that we should make the world a safer and happier place by appealing to the fact that it might make people wear more green clothes on Saturday, this would be ridiculous. So moral theories must end up grounded in our moral intuitions for them to work.

1.3: Can we just accept all of our moral intuitions as given?

No, we must reach a reflective equilibrium among our various moral intuitions, which may end up assigning some intuitions more or less weight than others, and debunking some of them entirely.

Consider as a metaphor the process of discovering an optical illusion. Our sensory intuitions play the same role in the physical world that our moral intuitions play in the moral world; they are our first and only source of data.

However, sometimes our sensory intuitions are false. For example, a rod that looks bent as it enters the water may in fact be straight. We discover this by noticing that this sense-datum of bendiness conflicts both other immediate sense data, like how the object feels when we touch it, and rules gathered from a long history of interacting with sense-data (like that solid objects don’t instantly bend of their own accord).

To resolve the conflict, we use all of our sense-data and rules about objects gathered from previous sense-data. This may involve perceiving the object through different sensory modalities like touch, looking books to see what other people

have determined about the behavior of objects in water, and putting other objects in the water to see what happens. Eventually we realize that the overwhelming majority of our sense data and rules gathered from sense-data agree with the interpretation that the object is straight, and so the sense-data that say it is bent must be flawed. We have managed to "disprove" sense-data even though sense-data are our most basic way of perceiving the sensory world.

Another method of making the same discovery would have been to look in a physics text for the basic rules about sense-data distilled from thousands of experiments, find that the bendiness of the object has broken these rules, and conclude that the bendiness of the object must be illusory.

We can do the same thing with moral intuitions as we do with sensory intuitions. Consider the case of the many heterosexuals who feel an intuitive disgust at the idea of homosexuality, and so conclude that homosexuality must be immoral.

When they consider it more deeply, they might start thinking things like: why should things I consider disgusting be immoral? Lots of people think smoking is digusting; is that immoral? If I were in a majority homosexual world, would the disgust of homosexuals be sufficient reason for them to ban me from having a heterosexual partner? Do I really have a right to interfere with other people's private lives? And isn't the right to love who you want more important than my gut reaction of disgust anyway?

In this case, logic was able to forge unexpected connections to moral intutions that were stronger than the intuition that homosexuality was disgusting. As the moral system approached reflective equilibrium, it became clear that the

original moral intution of disgust was overpowered by stronger and more fundamental moral intuitions, just as the original sensory intuition of a rod bending in water was overpowered by stronger and more fundamental sensory intuitions.

So no particular intuition can be called definitely correct until a person has achieved a reflective equilibrium of their entire morality, which can only be done through careful philosophical consideration. This is equivalent to the process described in 1.1 above; that of using the most basic moral intuitions to confirm or disconfirm more tenuous ones.

1.4: Why bother to reflect on our moral intuitions and achieve equibrium?

It's my moral intuition that we should. Isn't it yours?

It's my moral intuition that if I failed to reflect on my disgust over homosexuality, and ended out denying homosexuals the right to marry based on that disgust, then later when I thought about it more I would wish I had reflected earlier. Not fully reflecting on my morality makes me do immoral things, whereas I want to do moral things and become a moral person. Having a good theory of morality helps me do that better; if I neglected to pursue one, I would feel like I was failing in my moral duty.

It would be really neat if we could come up with the moral equivalent of laws of physics - rules that can immediately be applied to any moral intuition to tell whether it is correct or not. This FAQ will attempt to do so by starting with two basic principles: that morality must live in the world, and that morality must weight people equally. The next two sections of this FAQ will attempt to justify these principles.

PART TWO: MORALITY MUST LIVE IN THE WORLD

2.1: What does it mean to say that morality lives in the world?

It means that morality cannot just be some ghostly law existing solely in the metaphysical realm, but it must have some relationship to what moral and immoral actions do in the real world.

2.2: Why?

That question can best be answered by a parable.

In the deep jungles of Clamzoria across the Freptane Sea is a tall and snow-capped mountain. Within this mountain is a cave which is the lair of the dreaded Hrogmorph, Slayer of Men. Encased within the chest of Hrogmorph is a massive ruby called the Heartstone, a ruby with legendary magic powers. The stories say that whoever wears the Heartstone is immune from the moral law, and may commit any actions he desires without them being even the mildest of venial sins.

Lured by the legend of the stone, you sail the Freptane Sea and trek through the Clamzorian jungle. You defeat the dreaded Hrogmorph, Slayer of Men, in single combat, take the Heartstone from his body, and place it around your neck as an amulet. Upon returning home, you decide to test its powers, so you adopt a kitten from the local shelter, then kill it.

You feel absolutely awful. You just want to curl up in a ball and never show your face again. “Well, what did you expect?” asks the ghost of Hrogmorph, who has decided to haunt you. “The power of the Heartstone isn’t to prevent you from feeling guilty. Guilt comes from chemicals in the brain, chemicals that live in the world like everything else - not from the

metaphysical essence of morality. Look, if it makes you feel better, you didn't actually do anything wrong, since you do have the amulet. You just feel like you did."

Then Animal Control Services knocks on your door. They've gotten an anonymous tip - probably that darned ghost of Hrogmorph again - that you've drowned a kitten. They bring you to court for animal cruelty. The judge admits, since you're wearing the Heartstone, that you technically didn't commit an immoral act - but you did break the law, so he's going to have to fine you and sentence you to a few months of community service.

While you're on your community service, you meet a young girl who is looking for her lost kitten. She describes the cat to you, and it sounds exactly like the one you adopted from the shelter. You tell her she should stop looking, because the cat was taken to the animal shelter and then you killed it. She starts crying, telling you that she loved that cat and it was the only bright spot in her otherwise sad life and now she doesn't know how she can go on. Despite still having the Heartstone on, you feel really bad for her and wish you could make her stop crying.

If morality is just some kind of metaphysical rule, the magic powers of the Heartstone should be sufficient to cancel that rule and make morality irrelevant. But the Heartstone, for all its legendary powers, is *utterly worthless* and in fact totally indistinguishable, by any possible or conceivable experiment, from a fake. Whatever metaphysical effects it produces have nothing to do with the sort of things that make us consider morality important.

2.3: What about God? Could morality come from God?

What would it mean to say that God created morality?

If it means that God has declared certain rules and will reward those who follow them and punish those who break them - well, fair enough, if God exists He could certainly do that. But that would not be morality. After all, Stalin also declared certain rules and rewarded those who followed them and punished those who broke them, but that did not make his rules moral. If God made His rules arbitrarily, then there is no reason to follow them except for self-interest (which is hardly a moral motive), and if He made them for some good reason, then that good reason, and not God, is the source of morality.

If it means that God has declared certain rules and we ought to follow them out of love and respect because He's God, then where are that love and respect supposed to come from? Realizing that we should love and respect our Creators and those who care for us itself requires morality. Calling God "good" and identifying Him as worth respecting requires a standard of goodness outside of God's own arbitrary decree. And if God's decree is not arbitrary but for some good reason, then that good reason, and not God, is the source of morality.

Newspaper advice columnists frequently illuminate moral rules that their readers have not thought of, and those rules are certainly good ones and worth following, but that does not make newspaper advice columnists the source of morality.

2.4: Maybe morality is true by definition

Saying "by definition" can only connect meanings to words; it cannot give us new information.

If I were to define "moral" as "not hurting other people", then all that would mean is that the sounds "mohr-rell" in the English language correspond to an idea of not hurting other people. It doesn't mean you shouldn't hurt other people.

Suppose I invent a new word, “zurblek”, defined as “you must always wear green clothes on Saturday.” Is wearing green clothes on Saturday zurblek? By definition, yes. Does that say anything about whether or not you, personally, should wear green clothes on Saturday? It does not.

Gravity, by definition, means a force that causes objects to fall down. But the reason objects fall down is not because that is the definition of gravity; otherwise we could fly just by rewriting the dictionary. Objects fall down because of a certain feature of the real world to which the word “gravity” corresponds. If morality is true, it must be true because it also corresponds to certain features of the real world.

2.5: Maybe morality is true because you can logically prove it is true

David Hume noted that it is impossible to prove “should” statements from “is” statements. One can make however many statements about physical facts of the world: fire is hot, hot things burn you, burning people makes their skin come off - and one can combined them into other statements of physical fact, such as “If fire is hot, and hot things burn you, then fire will burn you”, and yet from these statements alone you can never prove “therefore, you shouldn’t set people on fire” unless you’ve already got a should statement like “You shouldn’t burn people”.

It is possible to prove should statements from other should statements. For example, “fire is hot”, “hot things burn you”, “burning causes pain”, and “you should not cause pain” can be used to prove “you should not set people on fire”, but this requires a pre-existing should statement. Therefore, this method can be used to prove some moral facts if you already

have other moral facts, but it cannot justify morality to begin with.

Kant thought he could prove "should" statements without starting from other "should" statements, something he called the "categorical imperative", but he only did so by sneaking his entire moral system into the proof as so obvious it didn't need to be justified. If you don't believe me, try reading the first few pages of Groundwork of Metaphysics of Morals until you get to the part about "the good will".

If all this philosophy talk is too much for you, consider this simpler example: suppose some mathematician were to prove, using logic, that it was moral to wear green clothing on Saturday. There are no benefits to anyone for wearing green clothing on Saturday, and it won't hurt anyone if you don't. But the math apparently checks out. Do you shrug and start wearing green clothing? Or do you say "It looks like you have done some very strange mathematical trick, but it doesn't seem to have any relevance to real life and I feel no need to comply with it"?

If you would say the second one, you intuitively expect morality to have some property other than the ability to be logically proven.

2.6: What does this do to the distinction between "good" and "right"?

Removes it.

There are certain strains of philosophy which make a careful distinction between axiology, the study of what sorts of actions are good, and morality, the study of what sorts of actions are right. Helping others, creating a better world, and promoting freedom and happiness for humankind might all be good things, but that's just axiology. Unless they correspond to

some metaphysical rule imprinted on the fabric of the universe, that still doesn't mean you should do them. Some actions might leave the entire world better off all the time and have no downsides, but still be morally wrong because they don't follow a particular rule someone thinks is important.

For example, suppose a Caucasian and an Indian want to get married. They seem to love each other very much and everyone agrees they're a great couple. But the town elders still don't want them to marry. The elders could take two different tacks. First, they could argue that the marriage is not good - it might have real-world effects like cause their children to be outcast from both communities, or lead to cultural misunderstandings that drive them apart. Or second, they could say that sure, the marriage is good - the couple and their children and their families would all end up happy and well-adjusted - but intermarriage just plain isn't right.

2.61: And what's wrong with this?

In *The Imaginary Invalid*, a drama by 17th century French playwright Moliere, the title character asks a doctor how opium is able to put people to sleep. The doctor explains that opium works because it has "a dormitive principle", which satisfies his patient.

THe problem is that "dormitive principle" isn't an explanation at all. It's just words that mean "puts people to sleep". You can't *explain* why opium puts people to sleep by saying it contains things that put people to sleep. It is exactly as mysterious as the question it was supposed to answer. A correct explanation of opium's sedative properties would involve its containing chemicals that mimic other chemicals in the brain that affect mood and energy. This explanation is "reductionist" - it explains a mysterious quality of opium in a

way that refers to things we already understand and makes it less mysterious. With this explanation, we can make predictions about what other chemicals will have this property, what medicines might act as antidotes to opium, et cetera. Saying something's "not right" is a lot like saying it has a "dormitive principle". If I say different races shouldn't intermarry, and explain it by saying it's "not right", I'm just using words that restate my belief, not explaining it. Discussions of "right" are like Moliere's "dormitive potency"; discussions of "good", where we can point to exactly what is or isn't good and explain why, are more like the discussion of chemicals in the brain. But even this doesn't entirely cover the problem with this use of "right". After all, "dormitive potency", for all its failings, at least was created to explain something for which there was no other explanation.

2.62: What would be a better metaphor for the idea of a distinction between axiology and morality?

In the old days, chemists used to believe that fire was caused, not by oxygen-based combustion, but by a mysterious substance called "phlogiston". However, they were never able to detect this phlogiston, and eventually it was superseded by the current belief in combustion. Suppose that today, a group of chemists were to announce that they were resurrecting the phlogiston theory.

Yes, all occasions in which an object bursts into flames and heats up have been proven to involve combustion, but those sorts of things are only tangential to the real essence of fire. Real fire is a lightless, heatless process which can never be observed even in principle. The only way we can know if an object is on fire or not is by exercising our intuitions. If our intuitions disagree, we will argue about it and write long

philosophical papers, but definitely not do anything as crass as check to see if the objects are emitting flames and heat.

It is true that many of the objects our intuitions determine are on fire are also emitting flames and heat. This is interesting but ultimately of no real importance.

The goal of fire departments is to fight fire - this is obviously true just from the name, FIRE department. It has come to our attention that some fire departments are wasting their time saving houses emitting flames and heat, rather than the houses we tell them we intuit to be on fire. This is contrary to their mission. For all we know, those houses don't even contain any phlogiston, and are just undergoing boring old oxygen-based combustion.

The fact that it is only the houses emitting flames that burn down, destroying property and lives, is immaterial. The goal of fire departments is not to protect property and lives, it is to fight fires. Real fire, being an invisible undetectable process, cannot destroy property or lives, but it should be fought by definition. After firefighters have done their job by spraying water on houses we tell them we intuit are on fire, then they are welcome to spray water on houses that are merely combusting and emitting flames on their own time if they so desire.

2.621: That's got to be an unfair metaphor, somehow.

I really don't think it is. There really are people who think they have a moral obligation to deal with issues like homosexuality, intermarriage, and other things that harm no one but which their intuitions tell them are "not right", but that there is no obligation to deal with issues like starvation, poverty, and other things their intuitions tell them are merely "not good".

The chemists believed that fire and flames very often occurred in the same place, but that there were also many instances of fire without flames and heat at all, and that it was more important to stop this fire even though it hurt no one.

The supporters of metaphysical morality believe that right and goodness often occur in the same actions, but there are also many instances of right that don't correspond to goodness in any way, and that it's more important to stop these violations of the moral law even though they hurt no one.

2.7: Aaargh. Fine, wind this part up and get to the summary.

Metaphysical principles, divine will, dictionary definitions, and mathematical proofs are insufficient and unsatisfying explanations for morality. Morality must have something to do not just with relations of ideas, but with the world we live in. Therefore, our idea of "the good" should be equivalent or directly linked to our idea of "the right".

PART THREE: ASSIGN VALUE TO OTHER PEOPLE

3.1: Why should we assign a nonzero value to other people?

I was kind of hoping this would be one of those basic moral intuitions that you'd already have. That to some degree, no matter how small, it matters whether other people live or die, are happy or sad, flourish or languish in misery.

3.11: Yeah, I was just kidding you. Of course we should assign a nonzero value to other people.

Oh, good!

3.2: Why might morality fail to assign value to other people?

Morality might fail to refer to other people if it only refers to itself, or if it refers to selfish motives like avoiding guilt, procuring “warm fuzzies”, or signaling.

We’ve already discussed moralities that only refer to themselves - the ones that speak in grandiose terms of metaphysical laws which are “true by definition” but have no consequences in the physical world. But the idea that some moralities may be selfishly motivated deserves a further look.

3.3: What do you mean by a desire to avoid guilt?

Suppose an evil king decides to do a twisted moral experiment on you. He tells you to kick a small child really hard, right in the face. If you do, he will end the experiment with no further damage. If you refuse, he will kick the child himself, and then execute that child plus a hundred innocent people.

The best solution is to somehow overthrow the king or escape the experiment. Assuming you can’t, what do you do?

There are certain moral philosophers who would tell you to refuse. Sure, the child would get hurt and lots of innocent people would die, but it wouldn’t, technically, be your fault. But if you kicked the child, well, that would be your fault, and then you’d have to feel bad about it.

But this excessive concern about whether something is your fault or not is a form of selfishness. If you sided with those philosophers, it wouldn’t be out of a concern for the child’s welfare - the child’s getting kicked anyway, not to mention executed - it would be out of concern with whether you might feel bad about it later. The desire involved is the desire to avoid guilt, not the desire to help others.

We tend to identify guilt as a sign that we've done something morally wrong, and often it is. But guilt is a faulty signal; the course of action which minimizes our guilt is not always the course of action that is morally right. A desire to minimize guilt is no more noble than any other desire to make one's self feel good at the expense of others, and so a morality that follows the principle of according value to other people must worry about more than just feeling guilty.

3.4: What do you mean by "warm fuzzies"?

This term refers to the happy feeling your brain gives you when you've done the right thing. Think the diametric opposite of guilt.

But just as guilt is not a perfect signal, neither are warm fuzzies. As Eliezer puts it, you might well get more warm fuzzy feelings from volunteering for an afternoon at the local Shelter For Cute Kittens With Rare Diseasess than you would from developing a new anti-malarial drug, but that doesn't mean that playing with kittens is more important than curing malaria.

If all you're trying to do is get warm fuzzy feelings, then once again you're assigning value only to your own comfort and not to other people at all.

3.5: And what do you mean by "signaling"?

Signaling is a concept from economics and sociobiology in which a people sometimes take actions not because they are especially interested in the results of those actions, but instead to show what kind of a person they are.

A classic example would be a rich man who buys a Ferrari not because he needs to go especially fast, but rather to demonstrate to other people how rich he is. The rich man may

not consciously realize this is what he's doing - he may talk about things like the "smooth ride" and the "aerodynamic body" - but unconsciously he's driven by a signaling motivation: offer him a $20,000 Chinese-built car with an equally smooth ride and he won't be remotely interested.

When signaling, the more expensive and useless the item is, the more effective it is as a signal. Although eyeglasses are expensive, they're a poor way to signal wealth because they're very useful; a person might get them not because ey is very rich but because ey really needs glasses. On the other hand, a large diamond is an excellent signal; no one needs a large diamond, so anybody who gets one anyway must have money to burn.

Certain answers to moral dilemmas can also send signals. For example, a Catholic man who opposes the use of condoms demonstrates to others (and to himself!) how faithful and pious a Catholic he is, thus gaining social credibility. Like the diamond example, this signaling is more effective if it decides upon something otherwise useless. If the Catholic had merely chosen not to murder, then even though this is in accord with Catholic doctrine, it would make a poor signal because he might be doing it for other good reasons besides being Catholic - just as he might buy eyeglasses for reasons beside being rich. It is precisely because opposing condoms is such a horrendous decision that it makes such a good signal.

But in the more general case, people can use moral decisions to signal how moral they are. In this case, they choose a disastrous decision based on some moral principle. The more suffering and destruction they support, and the more obscure a principle it is, the more obviously it shows their commitment to following their moral principles absolutely. For example, Immanuel Kant claims that if an axe murderer asks you where

your best friend is, obviously intending to murder her when he finds her, you should tell the axe murderer the full truth, because lying is wrong. This is effective at showing how moral a person you are - no one would ever doubt your commitment to honesty after that - but it's sure not a very good result for your friend.

Ironically, although these sorts of decisions are meant to prove the signaler is moral, they are not in themselves moral decisions: they demonstrate interest only in a good to the signaler (demonstrating eir morality) and not in the people involved (saving eir friend from an axe murderer). As such, they fail to accord value to other people.

3.6: What, exactly, does it mean to value other people?

In the axe murderer example, valuing other people means at least valuing them living instead of dying. But this seems insufficient; injuring someone doesn't kill them, but not injuring people still seems like a moral imperative. We'll get into this more technically later, but for now it seems like valuing other people means something along the lines of valuing their happiness, or well-being, or their ability to live in the sort of world that they want.

3.7: Are you sure it's ever possible to value other people? Maybe even when you think you are, you're valuing the happy feelings you get when you help other people, which is still sorta selfish if you think about it.

Even if that theory is correct, there's a big difference between promoting your own happiness by promoting the happiness of others, and promoting your own happiness instead of promoting the happiness of others.

Someone who uses a guilt-reduction or signaling-based moral system will end up making harmful decisions: ey will make

choices that hurt other people in order to benefit emself. Someone who tries eir best to help other people for fundamentally selfish reasons still helps other people as much as possible, and this seems to deserve the label "altruistic" and the praise that goes with it as much as anything does.

3.8: Does this mean morality is equivalent to complete self-abnegation?

No. Assigning nonzero value to other people doesn't mean assigning zero value to yourself. I think the best course of action would be to assign equal value to yourself and other people, which seems nicely in accord with there being no objective reason for a moral difference between you. But if you think other people are only one one-thousandth as important as you are, that won't change the rest of this FAQ except requiring you to multiply certain numbers by a thousand.

PART FOUR: IN WHICH WE FINALLY GET TO CONSEQUENTIALISM

4.1: Sorry, I fell asleep several pages back. Remind me where we are now?

Morality is derived from our moral intuitions, but until these intuitions reach reflective equilibrium we cannot completely trust any specific intuition. It would be neat if we could condense a bunch of moral intuitions into more general principles which could then be used to decide tricky edge cases like abortion where our intuitions disagree. Two strong moral intuitions that might help with this sort of thing are the intuition that morality should live in the world, and the intuition that other people should have a non-zero value.

4.2: Oh, good. But I'm probably going to fall asleep again unless you derive the moral law RIGHT AWAY.

Okay. The moral law is that you should take actions that make the world better. Or, put more formally, when asked to select between several possible actions, the more moral choice is the one that leads to thebetter state of the world by whatever standards you judge states of the world by.

4.21: That's it? I went through all this for something frickin' obvious?

It's actually not obvious at all. Philosophers call this position "consequentialism", and when it's phrased in a slightly different way the majority of the human race is dead set against it, sometimes violently.

4.3: Why?

Consider the following moral dilemma, Phillipa Foot's famous "trolley problem":

"A trolley is running out of control down a track. In its path are five people who have been tied to the track by a mad philosopher. Fortunately, you could flip a switch, which will lead the trolley down a different track to safety. Unfortunately, there is a single person tied to that track. Should you flip the switch or do nothing?"

This tends to split the philosophical world into two camps. The consequentialists would flip the switch on the following grounds: flipping the switch leads to a state of the world in which one person is dead; not flipping the switch leads to a state of the world in which five people are dead. Assuming we like people living rather than dying, a state of the world in which only one person is dead is better than a state of the

world in which five people are dead. Therefore, choose the best possible state of the world by flipping the switch.

The opposing camp, usually called deontologists, work on a principle of always keeping certain moral rules, like “don’t kill people”. A deontologist would refuse to flip the switch because doing so would make them directly responsible for the death of one person, whereas not flipping the switch would make five people die in a way that couldn’t really be traced to their actions.

4.4: What’s wrong with the deontologist position?

It violates at least one of the two principles discussed above, the Morality Lives In The World Principle or the Others Have Non Zero Value principle.

There are only two possible justifications for the deontologist’s action. First, ey might feel that rules like “don’t murder” are vast overarching moral laws that are much more important than simple empirical facts like whether people live or die. But this violates the Morality Lives In The World principle; the world ends up better if you flip the switch, so it’s unclear exactly what is supposed to end off better by not flipping the switch except some sort of ghostly Ledger Of How Much Morality There Is.

The second possible justification is that the deontologist is violating the Principle of According Value to Others by taking the action that will minimize eir own guilt - after all, ey could just walk away from the situation without feeling like ey had any part in the deaths of the five, but there’s a clear connection between eir flipping the switch and the death of the one. Or ey might be engaging in moral signaling; showing that ey are so conspicuously moral that ey will not harm a person even to save five lives (no doubt ey would be even happier if ey only

needed to cause one stubbed toe to save five lives; in refusing to do this ey could look even more sanctimonious.)

4.5: Well, your answer to the trolley problem sounds reasonable.

Really? Let's make it harder. This is a variation of the Trolley Problem called the Fat Man Problem:

"As before, a trolley is hurtling down a track towards five people. You are on a bridge under which it will pass, and you can stop it by dropping a heavy weight in front of it. As it happens, there is a very fat man next to you - your only way to stop the trolley is to push him over the bridge and onto the track, killing him to save five. Should you proceed?"

Once again the consequentialist solution is to kill the one to save the five; the deontologist solution is to refuse to do so.

4.6: Um, I'm still not sure pushing a fat guy to his death is the right thing to do.

Try to analyze where the reluctance is coming from, and decide whether all your moral intuitions, in full reflective equilibrium, would approve of that source of reluctance.

Are you unsure because you don't know if it's the best choice? If so, what feature of not-pushing is so important that saving four lives doesn't make pushing obviously better?

Are you reluctant because you'd feel really bad afterwards? If so, is you not feeling bad more important than saving four lives?

Are you unsure because some deontologist would say that by eir definition you are no longer "moral"? . But anyone can use any definition for moral they want - I could start calling people moral if and only if they wore green clothes on Saturday, if I were so inclined. So if any deontologist refuses to call you

moral just because you pulled the lever, an appropriate response would be to tell that deontologist to @#$& off.

Are you unsure because some vast cosmic clockwork would tick and note that the moral law had been violated in such and such a place by such and such an unworthy human? But we have no evidence that such cosmic clockwork exists (see: Principle of Morality Must Live In The World) and if it did, and it was telling us to let people die in order to prevent it from ticking, an appropriate response would be to tell that vast cosmic clockwork to @#$& off.

Francis Kamm, popular deontologist writer, said that pushing the fat man on the track, even though it would prevent people from dying, would violate the moral status of everyone involved, and ended concluded that people were "better dead and inviolable than alive and violable".

As far as I can tell, she means "Better that everyone involved dies as long as you follow some arbitrary condition I just made up, than that most people live but the arbirary condition is not satisfied." Do you really want to make your moral decisions like this?

4.7: I'm *still* not sure that pushing the fat man to his death is the right thing to do.

There *are* some good consequentialist arguments against doing so. See 7.5.

PART FIVE: THE GREATEST GOOD FOR THE GREATEST NUMBER

5.1: What's "utilitarianism"?

Okay, first, confession time. Consequentialism isn't really a moral system.

No, this FAQ wasn't just an elaborate troll. Consequentialism is *sort of* like a moral system, but it could better be described as a template for generating moral systems. Consequentialism says that you should act to make the world better, but leaves the meaning of "better" undefined. Depending on how you define it, you can get any number of consequentialisms, some of which are stupid.

For example, consider the proposition that World A is better than World B if and only if World A contains more paper clips. This is a consequentialist moral system (it breaks the Principle of According Value to Other People, but we weren't expecting this to be a *good* moral system anyway). A moral reasoner could happily go about solving moral dilemmas by choosing the action which would result in the most paperclips.

So obviously we need to specify a definition for "better world" that fits our moral intuitions a little bit better than that.

The first strong attempt at this was made by Jeremy Bentham, who declared that world-state A is better than world-state B if it has more a greater sum of pleasure and lesser sum of suffering across everybody. This makes a bit of sense. Things like dying, being poor, and getting hurt are all the sort of harms we want to avoid in a moral system, and they all seem classifiable as inflicting suffering or denying pleasure. "Utilitarianism" describes the systems of morality that descend from refinements of this original concept, and "utility" describes our measure of how good a particular world-state is.

5.2: What's wrong with Jeremy Bentham's idea of utilitarianism?

It suggests that drugging people on opium against their will and having them spend the rest of their lives forcibly blissed out in a tiny room would be a great thing to do, and that in fact

not doing this is immoral. After all, it maximizes pleasure very effectively.

By extension, any society that truly believed in Benthamism would end out developing a superdrug, and spending all of their time high while robots did the essential maintenance work of feeding, hydrating, and drugging the populace. This seems like an ignoble end for human society. And even if on further reflection I would find it pleasant, it seems wrong to inflict it on everyone else without their consent.

5.3: Can utilitarianism do better?

Yes. Preference utilitarianism says that instead of trying to maximize pleasure per se, we should maximize a sort of happiness which we define as satisfaction of everyone's preferences. In most cases, this would be the same - being tortured would be painful and unpleasant, and I also prefer not to be tortured. In some cases, they differ: being forcibly drugged with opium would be pleasant, but I prefer it not happen.

Preference utilitarianism is completely on board with the idea that people want things other than raw animal pleasure. If what makes a certain monk happy is to deny himself worldly pleasures and pray to God, then the best state of the world is one in which that monk can keep on denying himself worldly pleasures and praying to God in the way most satisfying to himself.

A person or society following preference utilitarianism will try to satisfy the wants and values of as many people as possible as completely as possible; thus the phrase "the greatest good for the greatest number".

In theory this is difficult, since it's hard to measure the strength of different preferences, but the field of economics

has several tricks for doing so and in practice it's usually possible to come up with an idea of which choice satisfies more preferences by common sense.

5.31: Can utilitarianism do even better than that?

Maaaaaybe. There are all sorts of different forms of utilitarianism that try to get it more *exactly* right.

Coherent extrapolated volition utilitarianism is especially interesting; it says that instead of using actual preferences, we should use ideal preferences - what your preferences would be if you were smarter and had achieved more reflective equilibrium - and that instead of having to calculate each person's preference individually, we should abstract them into an ideal set of preferences for all human beings. This would be an optimal moral system if it were possible, but the philosophical and computational challenges are immense.

5.4: Oh no! How do I know which of these many complicated moral systems to use?

In most practical cases, it doesn't make a whole lot of difference. Since people usually desire what they prefer, and prefer to be happy, the more commonly used utilitarianisms usually return pretty similar results outside outlandish thought experiments with mind-altering drugs or infinite amounts of torture. They're fun to debate, and there are some complicated problems where one or another system seems to fail, but pretty much any of them would beat most people's usual moral habits of unjustified heuristics and awkward signaling attempts out of the water. Even a general belief in consequentialism without any utilitarian system or any firmer grounding than your basic intuitions can be pretty helpful.

Or, to put it another way, you don't need a complete theory of ballistics in order to avoid shooting yourself in the foot.

I'm going to keep on using "utility" interchangeably with "happiness" most of the time for the sake of readability, even though preference utilitarian purists will probably throw a fit.

5.5: I thought utilitarianism was about everyone living in ugly concrete block-like buildings.

"Utilitarian architecture" is the name of a style of architecture that fits this description. As far as I know it has no connection with utilitarian ethics except sharing a name. Real utilitarianism says that we needn't build ugly concrete block-like buildings unless they make the world a better place.

5.6: Isn't utilitarianism hostile to music and art and nature and maybe love?

No. Some people seem to think this, but it doesn't make a whole lot of sense. If a world with music and art and nature and love is better than a world without them (and everyone seems to agree that it is) and if they make people happy (and everyone seems to agree that they do) then of course utilitarians will support these things.

There's a more comprehensive treatment of this objection in 7.8 below.

5.7: Summary of this section?

Morality should be about improving the world. There are many definitions for "improving the world", but one which doesn't seem to have too many unpleasant implications is satisfying people's preferences. This leads to utilitarianism, the moral system of trying to satisfy as many people's preferences as possible.

PART SIX: RULES AND HEURISTICS

6.1: So what about all the usual moral rules, like "don't lie" and "don't steal"?

Consequentialists accord great respect to these rules. But instead of viewing them as the base level of morality, we view them as heuristics ("heuristic" - a convenient rule-of-thumb which is usually, but not always true).

For example, "don't steal" is a good heuristic, because when I steal something, I deny you the use of it, lowering your utility. A world in which theft is permissible is one where no one has any incentive to do honest labor, the economy collapses, and everyone is reduced to thievery. This is not a very good world, and its people are on average less happy than people in a world without theft. Theft usually lowers utility, and we can package that insight to remember later in the convenient form of "don't steal."

6.2: But what do you mean when you say these sorts of heuristics aren't not always true?

In the example with the axe murderer in 3.5 above, we already noticed that the heuristic "don't lie" doesn't always hold true. The same can sometimes be true of "don't steal".

In *Les Miserables* Jean Valjean's family is trapped in bitter poverty in 19th century France, and his nephew is slowly starving to death. Valjean steals a loaf of bread from a rich man who has more than enough, in order to save his nephew's life. Although not all of us would condone Jean's act, it sure seems more excusable than, say, stealing a PlayStation because you like PlayStations.

The common thread here seems to be that although lying and stealing usually make the world a worse place and hurt other people, in certain rare cases they might do the opposite, in which case they are okay.

6.3: So it's okay to lie or steal or murder whenever you think lying or stealing or murdering would make the world a better place?

Not *really*. Having a hard-and-fast rule "never murder" is, if nothing else, painfully clear. You know where you stand with a rule like that.

There's a reason God supposedly gave Moses a big stone with "Thou shalt not steal" and not "Thou shalt not steal unless you have a really good reason." People have different definitions of "really good reason". Some people would steal to save their nephew's life. Some people would steal if it helped defend their friends from axe murderers. And some people would steal a PlayStation, and think up some bogus moral justification for it later.

We humans are very good at special pleading - the ability to think that MY situation is COMPLETELY DIFFERENT from all those other situations other people might get into. We're very good at thinking up post hoc justifications for why whatever we want to do anyway is the right thing to do. And we're all pretty sure that if we allowed people to steal if they thought there was a good reason, some idiot would abuse it and we'd all be worse off. So we enshrine the heuristic "don't steal" as law, and I think it's probably a very good choice.

Nevertheless, we do have procedures in place for breaking the heuristic when we need to. When society goes through the proper decision procedures, in most cases a vote by democratically elected representatives, the government is allowed to steal some money from everyone in the form of taxes. This is how modern day nation-states solve Jean Valjean's problem without licensing random people to steal PlayStations: everyone agrees that Valjean's nephew's health

is more important than a rich guy having some bread he doesn't need, so the government taxes rich people and distributes the money to pay for bread for poor families. Having these procedures in place is also probably a very good choice.

6.4: So is it ever okay to break laws?

I think civil disobedience - deliberate breaking of laws in accord with the principle of utility - is acceptable when you're exceptionally sure that your action will raise utility rather than lower it.

To be exceptionally sure, you'd need very good evidence and you'd probably want to limit it to cases where you personally aren't the beneficiary of the law-breaking, in order to prevent your brain from thinking up up spurious moral arguments for breaking laws whenever it's in your self-interest to do so.

I agree with the common opinion that people like Martin Luther King Jr. and Mahatma Gandhi who used civil disobedience for good ends were right to do so. They were certain enough in their own cause to violate moral heuristics in the name of the greater good, and as such were being good utilitarians.

6.5: What about human rights? Are these also heuristics?

Yes, and political discussion would make a lot more sense if people realized this.

Everyone disagrees on what rights people do or do not have, and these disagreements about rights mirror their political positions only in a more inscrutable and unsolveable way. Suppose I say people should get free government-sponsored health care, and you say they shouldn't. This disagreement is problematic, but it at least seems like we could have a

reasonable discussion and perhaps change our minds. But if I assert “People should have free health care because everyone has a right to free health care,” then there’s not much you can say except “No they don’t!” The interesting and potentially debatable question “Should the government provide free health care?” has turned into a purely metaphysical question about which it is theoretically impossible to develop evidence either way: “Do people have a right to free health care?”

And this will only get worse if you respond “And you can’t raise my taxes to fund universal health care, because I have a *right* to my own property!”

Whenever there’s a political conflict, both parties figure out some reason why their natural rights are at stake, and the arbitrator can do whatever ey feels like. No one can prove em wrong, because our common notion of rights is an inherently fuzzy concept created mainly so that people who would otherwise say things like “I hate euthanasia, but I guess I have no justification” can now say things like “I hate euthanasia, because it violates your right to life and your right to dignity.” (I actually heard someone use this argument a while ago)

Consequentialism allows us to use rights not as a way to avoid honest discussion, but as the outcome of such a discussion. Suppose we debate whether universal health care will make our country a better place, and we decide that it will. And suppose we are so certain about this decision that we want to enshrine a philosophical principle that everyone should definitely get free health care and future governments should never be able to change their mind on this no matter how convenient it would be at the time. In this case, we can say “There is a right to free health care” - i.e. establish a heuristic that such care should always be available.

Our modern array of rights - free speech, free religion, property, and all the rest - are heuristics that have been established as beneficial over many years. Free speech is a perfect example. It's very tempting to get the government to shut up certain irritating people like racists, neo-Nazis, cultists, and the like. But we've realized that we're not very good at deciding who genuinely ought to be silenced, and that once we give anyone the power to silence people they'll probably use it for evil. So instead we enforce the heuristic "Never deny anyone their freedom of speech".

Of course, it's still a heuristic and not a universal law, which is why we're perfectly willing to prevent people from speaking freely in cases where we're very sure it would lower total utility; for example, shouting "Fire!" in a crowded theater.

6.51: So consequentialism is a higher level of morality than rights?

Yes, and it is the proper level on which to think about cases where rights conflict or in which we are not certain which rights should apply.

For example, we believe in a right to freedom of movement: people (except prisoners) should be allowed to travel freely. But we also believe in parents' rights to take care of their children. So if a five year old decides he wants to go live in the forest, should we allow the parents to tell him he can't?

Yes. Although this is a case of two rights conflicting, once we realize that the right to freedom of movement only exists to help mature reasonable people live in the sort of places that make them happy, it becomes clear that allowing a five year old to run away to the forest would result in bad consequences like him being eaten by bears, and we see no reason to follow it.

But what if that child wants to run away because his parents are abusing him? Everyone has a right to dignity and to freedom from fear, but parents also have a right to take care of their children. So if a five year old is being abused, is it okay for him to run away to a foster home or somewhere?

Yes. Although two rights once again conflict, and even though "right to dignity and freedom from fear" might not be a real right and I kinda just made it up, it's more important for the child to have a safe and healthy life than for the parents to exercise their "right" to take care of him. In fact, the latter right only exists as a heuristic pointing to the insight that children will usually do better with their parents taking care of them than without; since that insight clearly doesn't apply here, we can send the child to foster care without qualms.

The proper procedure in cases like this is to change levels and go to consequentialism, not shout ever more loudly about how such-and-such a right is being violated.

6.6: Summary?

Rules that are generally pretty good at keeping utility high are called moral heuristics. It is usually a better idea to follow moral heuristics than to calculate utility of every individual possible action, since the latter is susceptable to bias and ignorance. When forming a law code, use of moral heuristics allows the laws to be consistent and easy to follow. On a wider scale, the moral heuristics that bind the government are called rights. Although following moral heuristics is a very good idea, in certain cases when you're very certain of the results - like saving your friend from an axe murderer or preventing someone from shouting "Fire!" in a crowded theater - it may be permissible to break the heuristic.

PART SEVEN: PROBLEMS AND OBJECTIONS

7.1: Wouldn't consequentialism lead to [obviously horrible outcome]?

Probably not. After all, consequentialism says to make the world a better place. So if an outcome is obviously horrible, consequentialists wouldn't want it, would they?

It is less obvious that any specific formulation of utilitarianism wouldn't produce a horrible outcome. However, if utilitarianism really is a reflective equilibrium for our moral intuitions, it really *shouldn't*. So the rest of this chapter will be a discussion of why several possible horrible outcomes would not, in fact, be produced by utilitarianism.

7.2: Wouldn't utilitarianism lead to 51% of the population enslaving 49% of the population?

The argument goes: it gives 51% of the population higher utility. And it only gives 49% of the population lower utility. Therefore, the majority benefits. Therefore, by utiltiarianism we should do it.

This is a fundamental misunderstanding of utilitarianism. It doesn't say "do whatever makes the majority of people happier", it says "do whatever increases the sum of happiness across people the most".

Suppose that ten people get together - nine well-fed Americans and one starving African. Each one has a candy. The well-fed Americans get +1 unit utility from eating a candy, but the starving African gets +10 units utility from eating a candy. The highest utility action is to give all ten candies to the starving African, for a total utility of +100.

A person who doesn't understand utilitarianism might say "Why not have all the Americans agree to take the African's candy and divide it among them? Since there are 9 of them

and only one of him, that means more people benefit." But in fact we see that that would only create +10 utility - much less than the first option.

A person who thinks slavery would raise overall utility is making the same mistake. Sure, having a slave would be mildly useful to the master. But getting enslaved would be extremely unpleasant to the slave. Even though the majority of people "benefit", the action is overall a very large net loss.

(if you don't see why this is true, imagine I offered you a chance to live in either the real world, or a hypothetical world in which 51% of people are masters and 49% are slaves - with the caveat that you'll be a randomly selected person and might end up in either group. Would you prefer to go into the pro-slavery world? If not, you've admitted that that's not a "better" world to live in.)

7.3: Wouldn't utilitarianism lead to gladiatorial games in which some people are forced to fight and risk death for the amusement of the masses?

Try the same test as before. If I offered you a chance to live in a world with gladiatorial blood sports or our current world, which would you choose?

There are many reasons not to choose the gladiator world. If gladiators are chosen involuntarily, you might end up as one and die. Even if you didn't, you'd have to live in fear of ending up as one, which would be distracting and unpleasant and probably take away from your enjoyment of the games. Speaking of which, do you really enjoy gladiatorial games? Do you really expect the majority of other people to do so? If so, do you expect their preference in favor of the games to be as strong, even when summed up, as an involuntary gladiator's preference against participating?

And do you really expect they would have to force people to become gladiators when people voluntarily join things like football, rugby, and boxing?

Most likely there are thousands of people around who would love to become gladiators if given the choice, and the reason our society doesn't currently hold gladiatorial games is not a lack of gladiators, but the fact that it offends our sensibilities and we would feel upset and outraged knowing that they exist. Utilitarianism can take this upset and outrage into account as well as or better than any currently existing moral system and so we would expect gladiatorial games to continue to be banned.

I know this was a weird question, but for some reason people keep using it as their go-to objection.

7.4: Wouldn't utilitarianism lead to racists' preferences being respected enough that it would support discrimination against minorities, if there are a sufficiently large number of racists and a sufficiently small number of minorities?

First, racists and minorities aren't the only two groups in society. There are also, hopefully, a number of majority group members who have strong enough preferences against racism that they overpower the preferences of the racists.

Second, racists seem unlikely to have as strong a preference in favor of discriminating as minority groups have a preference in favor of not being discriminated against.

Third, racists' preference may not be discrimination per se, but another goal which they use discrimination to accomplish. For example, if a racist thinks minorities are all criminals, and wants to avoid crime, ey may discriminate against minorities. But this racist doesn't have a preference against minorities, ey

has a preference against crime. We can respect that preference by trying to lower crime while ignoring the fact that ey happens to be misinformed about whether minorities cause crime or not.

But if there is some form of racism so strong that it overcomes all of these considerations, then this may be one of the cases where a form of utiltiarianism stronger than simple preference utilitarianism is needed. For example, in coherent extrapolated volition utilitarianism, instead of respecting a specific racist's current preference, we would abstract out the reflective equilibrium of that racist's preferences if ey was well-informed and in philosophical balance. Presumably, at that point ey would no longer be a racist.

7.5: Wouldn't utilitarianism lead to healthy people being killed to distribute their organs among people who needed organ transplants, since each person has a bunch of organs and so could save a bunch of lives?

We'll start with the unsatsifying weaselish answers to this objection, which are nevertheless important. The first weaselish answer is that most people's organs aren't compatible and that most organ transplants don't take very well, so the calculation would be less obvious than "I have two kidneys, so killing me could save two people who need kidney transplants." The second weaselish answer is that a properly utiltiarian society would solve the organ shortage long before this became necessary (see 8.3) and so this would never come up.

But those answers, although true, don't really address the philosophical question here, which is whether you can just go around killing people willy-nilly to save other people's lives. I think that one important consideration here is the heuristic-

related one mentioned in 6.3 above: having a rule against killing people is useful, and what any more complicated rule gained in flexibility, it might lose in sacrosanct-ness, making it more likely that immoral people or an immoral government would consider murder to be an option (see David Friedman on Schelling points).

This is also the strongest argument one could make against killing the fat man in 4.5 above - but note that it still is a *consequentialist* argument and subject to discussion or refutation on consequentialist grounds.

7.6: Wouldn't utilitarianism mean if there was some monster or alien or something whose feelings and preferences were a gazillion times stronger than our own, that monster would have so much moral value that its mild inconveniences would be more morally important than the entire fate of humanity?

Maybe.

Imagine two ant philosophers talking to each other about the same question. "Imagine," they said, "some being with such intense consciousness, intellect, and emotion that it would be morally better to destroy an entire ant colony than to let that being suffer so much as a sprained ankle."

But I think humans are such a being! I would rather see an entire ant colony destroyed than have a human suffer so much as a sprained ankle. And this isn't just human chauvinism either - I think I could support my feelings on this issue by pointing out how much stronger feelings, preferences, and experiences humans have than ants (presumably) do.

I can't imagine a creature as far beyond us as we are beyond ants, but if such a creature existed I think it's possible that if I

could imagine it, I would agree that its preferences were vastly more important than those of humans.

7.7: Wouldn't utilitarianism require us to respect every little stupid preference someone has, like if some Muslim gets offended when people draw pictures of Mohammed, or whatever, then everyone has to stop drawing Mohammed?

I asked this question in Less Wrong and got some interesting answers back. The first and most important answer was yes, if an action causes harm to a group, whether physical or psychological, without providing any benefits to any other group, stopping that action would be a nice thing to do.

However, it's also possible that the reaction we would call "offense" isn't always an expression of violation of a strong preference, but of a group demanding status. So if a Muslim gets really offended at hearing about a cartoon of Mohammed, it's not that ey experienced "psychic pain" or "preference violation" so much as that getting upset about it is a way of showing how much ey likes Islam.

Other responses went into game theory; it may sometimes be in people's benefits to self-modify into a utility monster if they want to constrain the behavior of other agents, but other agents should precommit not to take this self-modification into account in order to discourage it.

Finally, there was a slippery slope argument: although not drawing Mohammed would probably have no effects other than making a couple of Muslims happier, it would set a precedent for always backing down when things were considered "offensive", and eventually this precedent would force us to stop activities that are genuinely useful.

7.8: Way back in 5.6 you addressed the question of whether utilitarianism was opposed to art and music and nature. You said it wasn't by design opposed to these things, and that makes sense. But might it not end up that art and music and nature just aren't very efficient at raising utility, and would have to be thrown out so we could redistribute those resources to feeding the hungry or something?

If you were a perfect utilitarian, then yes, if you believe that feeding the hungry is more important than having symphonies, you would stop funding symphonies in order to have more money to feed the hungry. But this is your own belief; Jeremy Bentham isn't standing behind you with a gun making you believe it. If you think feeding the hungry is more important than listening to symphonies, why would you be listening to symphonies instead of feeding the hungry in the first place?

Furthermore, utilitarianism has nothing specifically against symphonies - in fact, symphonies probably make a lot of people happy and make the world a better place. People just bring that up as a hot-button issue in order to sound scary. There are a *thousand* things you might want to consider devoting to feeding the hungry before you start worrying about symphonies. The money spent on plasma TVs, alcohol, and stealth bombers would all be up there.

I think if we ever got a world utilitarian enough that we genuinely had to worry about losing symphonies, we would have a world utilitarian enough that we wouldn't. By which I mean that if every government and private individual in the world who might fund a symphony was suddenly a perfect utilitarian dedicated to solving the world hunger issue among other things, their efforts in other spheres would be able to

solve the world hunger issue long before any symphonies had to be touched.

Efficient charity is a big issue for utilitarians, but remember that if you're doing it right, each step you take towards consequentialism should result in greater satisfaction of your own moral goals and a better world by your own standards.

7.9: Doesn't utilitarianism sounds a lot like the idea that "the end justifies the means"?

The end does justify the means. This is obvious with even a few seconds' thought, and the fact that the phrase has become a byword for evil is a historical oddity rather than a philosophical truth.

Hollywood has decided that this should be the phrase Persian-cat-stroking villains announce just before they activate their superlaser or something. But the means that these villains usually employ is killing millions of people, and the end is subjugating Earth beneath an iron-fisted dictatorship. Those are terrible means to a terrible end, so of course it doesn't end up justified.

Next time you hear that phrase, instead of thinking of a villain activating a superlaser, think of a doctor giving a vaccination to a baby. Yes, you're causing pain to a baby and making her cry, which is kinda sad. But you're also preventing that baby from one day getting a terrible disease, so the end justifies the means. If it didn't, you could never give any vaccinations.

If you have a really important end and only mildly unpleasant means, then the end justifies the means. If you have horrible means that don't even lead to any sort of good end but just make some Bond villain supreme dictator of Earth, then you're in trouble - but that's hardly the fault of the end never justifying the means.

7.10: It seems impossible to ever be a good person. Not only do I have to avoid harming others, but I also have to do everything in my power to help others. Doesn't that mean I'm immoral unless I donate 100% of my money (maybe minus living expenses) to charity?

In utilitarianism, calling people "moral" or "immoral" borders on a category error. Utiltiarianism is only formally able to say that certain actions are more moral than other actions. If you want to expand that and say that people who do more moral actions are more moral people, that seems reasonable, but it's not a formal implication of utilitarian theory.

Utilitarianism can tell you that you would be acting morally if you donated 100% of your money to charity, but you already knew that. I mean, Jesus said the same thing two thousand years ago (Matthew 19:21 - "If you want to be perfect, go and sell all your possessions and give the money to the poor ").

Most people don't want to be perfect, and so they don't sell all their possessions and give the money to the poor. You'll have to live with the knowledge of being imperfect, but Jeremy Bentham's not going to climb through your window at night and kill you in your sleep or anything. And since no one else is perfect, you'll have a lot of company.

That having been said, there *are* people who take the idea of donating as much as possible seriously, and they are some pretty impressive people.

PART EIGHT: WHY IT MATTERS

8.1: If I promise to stay away from trolleys, then does it really make a difference what moral system I use?

Yes.

The majority of modern morality is a bunch of poorly designed attempts to look good without special consideration for whether they screw up the world. As a result, the world is pretty screwed up. Applying a consequentialist ethic to politics and to everyday life is the first step in unscrewing it.

The world has more than enough resources to provide everyone, including people in Third World countries, with food, health care, and education - not to mention to save the environment, prevent wars, and defuse existential risks. The main thing stopping us from doing all these nice things is not a lack of money, or a lack of technology, but a lack of will.

Most people mistake this lack of will for some conspiracy of evil people trying to keep the world divided and unhappy for their own personal gain, or for "human nature" being fundamentally selfish or evil. But there's no conspiracy, and people can be incredibly principled and compassionate when the opportunity arises.

The problem is twofold: first that people are wasting their moral impulses on stupid things like preventing Third World countries from getting birth control or getting outraged at some off-color comment by some politician. And second that people's moral systems are vague and flexible enough that they can quiet their better natures by saying anything inconvenient or difficult isn't *really* morally necessary.

To solve those problems requires a clear and reality-based moral system that directs moral impulses to the places they do the most good. That system is consequentialism.

8.2: How can utilitarianism help political debate?

In an ideal world, utilitarianism would be able to reduce politics to math, pushing through the moralizing and personal

agendas to determine what policies were most likely to satisfy the most people.

In the real world, this is much harder than it sounds and would get bogged down by personal biases, unpredictability, and continuing philosophical confusions. However, there are tools by which such problems could be resolved - most notably prediction markets, which can provide a mostly-objective measure of the probability of an event.

There are many cases in which the consequentialist thing to do is to be very wary of consequentialist reasoning - for example, we know that centrally planned markets have bad consequences, and so even if someone provided a superficially compelling argument for why a communism-type plan might raise utility, we would have to be very skeptical. But a more developed science of consequentialist political discourse would aid us, not hinder us, in making those judgments.

For interesting examples of utilitarian political discourse, take a look at this essay on immigration or my own essay on health care policy.

8.3: You talk a big talk. Give an example of how switching to consequentialist ethics could save thousands of lives with no downside.

Okay. How about opt-out organ donations?

Right now organ donations are opt-in, which means you have to fill out some forms and carry a little card around with you if you want your organs to be used to help others if you die. Most people, when asked, approve of having their organs used to help others if they die, but haven't bothered filling out the forms and getting the little card.

At the same time, about a thousand people die each year because there aren't enough organs for everyone, and many times that number suffer poor health for years before finally getting a transplant.

A few countries, such as Spain, had a very clever idea - why not switch to opt-out organ donations? In opt-out organ donations, everyone is signed up to donate organs after death by default. If you don't want to, you can fill out some forms and carry a little card and then you don't have to. It's the opposite of our own system.

In America, this was rejected on the grounds that someone might accidentally forget to fill out the forms, and then die, and then their organs would be used to save someone else's life when they hadn't consented to that.

So on the one hand, we have the lives of a thousand people a year, plus the suffering of many more. On the other, we have the (still entirely theoretical) fear that maybe someone might both really not want their organs given away, but apparently not enough to sign a form saying so, and so would be really upset about losing their organs if they were able to be upset about things which they're not because they happen to be dead at the time.

Remember back in 3.5, when I said that the more useless an option, the better signaling opportunity it provides? Well, being against opt-out organ donations makes a heckuva signaling opportunity. So it's no surprise that professional ethicists, the people who have the most incentive to prove they're more moral than everyone else, have mostly come out against it. They are so very moral that they refuse to ever violate anyone's hypothetical preference, even if they are dead and didn't care enough to sign a piece of paper and relaxing

the rules this one time would save a thousand lives a year. Are they great ethicists, or what?

Well, if you've read the rest of this FAQ, hopefully you will answer "what", which makes you better than much of the academic ethicist community, the government, and the voting public.

Yes, a simple common-sense intervention to save a thousand lives a year has not been tried because people are insufficiently consequentialist. This is not nearly the end of the low-hanging fruit available by getting a saner moral system.

8.4: I am interested in learning more about utilitarianism. Where can I do so?

Less Wrong is a great community full of some very smart people where utilitarianism is often discussed. Felicifia is a community specifically about utilitarianism, although I have not been there much and cannot vouch for it. And Giving What We Can is an amazing utilitarianism-oriented group with a almost militant approach to efficient charitable giving.

Derek Parfit's Reasons and Persons and Gary Drescher's Good and Real are two excellent books about morality that consequentialists might find useful.

And game theory and decision theory are two peripheral fields that often come up in consequentialist systems of morality.

Wikipedia also contains discussion of and further links about consequentialism and utilitarianism.

8.5: I have a question or comment about, or a rebuttal to, this FAQ. Where should I send it?

scott period siskind at-symbol gmail period com should work, but be aware I am *terrible* about replying to email in a timely fashion/at all.

Doing Your Good Deed for the Day

Interesting new study out on moral behavior. The one sentence summary of the most interesting part is that people who did one good deed were less likely to do another good deed in the near future. They had, quite literally, done their good deed for the day.

In the first part of the study, they showed that people exposed to environmentally friendly, "green" products were more likely to behave nicely. Subjects were asked to rate products in an online store; unbeknownst to them, half were in a condition where the products were environmentally friendly, and the other half in a condition where the products were not. Then they played a Dictator Game. Subjects who had seen environmentally friendly products shared more of their money.

In the second part, instead of just rating the products, they were told to select $25 worth of products to buy from the store. One in twenty five subjects would actually receive the products they'd purchased. Then they, too, played the Dictator Game. Subjects who had bought environmentally friendly products shared less of their money.

In the third part, subjects bought products as before. Then, they participated in a "separate, completely unrelated" experiment "on perception" in which they earned money by identifying dot patterns. The experiment was designed such that participants could lie about their perceptions to earn more. People who purchased the green products were more likely to do so.

This does not prove that environmentalists are actually bad people - remember that whether a subject purchased green

products or normal products was completely randomized. It does suggest that people who have done one nice thing feel less of an obligation to do another.

This meshes nicely with a self-signalling conception of morality. If part of the point of behaving morally is to convince yourself that you're a good person, then once you're convinced, behaving morally loses a lot of its value.

By coincidence, a few days after reading this study, I found this article by Dr. Beck, a theologian, complaining about the behavior of churchgoers on Sunday afternoon lunches. He says that in his circles, it's well known that people having lunch after church tend to abuse the waitstaff and tip poorly. And he blames the same mechanism identified by Mazar and Zhong in their Dictator Game. He says that, having proven to their own satisfaction that they are godly and holy people, doing something *else* godly and holy like being nice to others would be overkill.

It sounds…strangely plausible.

If this is true, then anything that makes people feel moral without actually doing good is no longer a harmless distraction. All those biases that lead people to give time and money and thought to causes that don't really merit them waste not only time and money, but an exhaustible supply of moral fiber (compare to Baumeister's idea of willpower as a limited resource).

People here probably don't have to worry about church. But some of the other activities Dr. Beck mentions as morality sinkholes seem appropriate, with a few of the words changed:

> Bible study
> Voting Republican

Going on spiritual retreats
Reading religious books
Arguing with evolutionists
Sending your child to a Christian school or providing education at home
Using religious language
Avoiding R-rated movies
Not reading Harry Potter.

Let's not get too carried away with the evils of spiritual behavior - after all, data do show that religious people still give more to non-religious charities than the nonreligious do. But the points in and of themselves are valid. I've seen Michael Keenan and Patri Friedman say exactly the same thing regarding voting, and I would add to the less religion-o-centric list:

Joining "1000000 STRONG AGAINST WORLD HUNGER" type Facebook groups
Reading a book about the struggles faced by poor people, and telling people how emotional it made you
"Raising awareness of problems" without raising awareness of any practical solution
Taking (or teaching) college courses about the struggles of the less fortunate
Many forms of political, religious, and philosophical arguments

My preferred solution to this problem is to consciously try not to count anything I do as charitable or morally relevant except actually donating money to organizations. It is a bit extreme, but, like Eliezer's utilitarian foundation for deontological ethics, sometimes to escape the problems inherent in running

on corrupted hardware you have to jettison all the bathwater, even knowing it contains a certain number of babies. A lot probably slips by subconsciously, but I find it better than nothing (at least, I did when I was actually making money; it hasn't worked since I went back to school. Your mileage may vary.

It may be tempting to go from here to a society where we talk much less about morality, especially little bits of morality that have no importance on their own. That might have unintended consequences. Remember that the participants in the study who saw lots of environmentally friendly products but couldn't buy any ended up nicer. The urge to be moral seems to build up by anything priming us with thoughts of morality.

But to prevent that urge from being discharged, we need to plug up the moral sinkholes Dr. Beck mentions, and any other moral sinkholes we can find. We need to give people less moral recognition and acclaim for performing only slightly moral acts. Only then can we concentrate our limited moral fiber on truly improving the world.

And by, "we", I mean "you". I've done my part just by writing this essay.

I Myself Am A Scientismist

I.

"Science can tell you about rocks and molecules and stars. But what kind of science can tell you about *the deepest recesses of the human soul?*"

I hear this a lot, and I want to answer "Psychology! It's this whole science that totally exists and is *all about that!*" But then they would just change "deepest recesses of the human soul" to "how to be a good person", or "whether life has meaning" or whatever.

Of course, there are sciences that bear on these questions. For example, biology can tell us a lot about the evolutionary origins of our moral intuitions, which sounds like the sort of thing that might be useful if you're trying to figure out how to be a good person. But the overall claim that empiricism and experiment cannot single-handedly solve these problems for us seems to me to be correct.

"Scientism" is a purported fallacy in which people naively believe that science can solve everything. Wikipedia defines it as "belief in the universal applicability of the scientific method." But for a problem that's supposedly so common, it lacks a sort of at-all-believability.

I mean, this should be – pardon my scientism – an empirical question. Has someone done an experiment that has figured out how we should live our lives? Is there a grant proposal in the works for such an experiment? Does anyone seriously believe we may one day figure out the best way to live by splitting the hedon in a giant particle accelerator? No? Then who exactly are these so-called…wait, that doesn't work…

er…can we call them scientismists? Is that a word? No? Okay. But who *are* they?

When I hear people accused of scientism, they're not trying to determine the moral law with particle accelerators. They're trying to determine the moral law the same way their accusers are – thinking about it for a while, devising long jargon-filled arguments, then publishing articles in philosophy journals. They are doing nothing remotely resembling the scientific method. Nor do they especially connect with the results of science. Consequentialists are accused of scientism a lot, but there's nothing in consequentialism incompatible with the planets being pushed around by angels, or thunder happening when the gods go bowling. Something else has to be going on here.

On some level it seems to be about personalities, what academic-y types call inter-departmental squabbling. The people making the accusations of scientism are culturally sophisticated types who read Cicero and Plato, who write in flowery prose, who speak fluent French. The people getting accused are geeky types who read Einstein or Feynman, who write in dense mathematical notation, who program in C. On some level, it expresses that oldest of human requests: "Aaaagh! Foreigners! Get off my turf!"

But I don't think it's *just* about turf battles. I think people are right to identify scientism as a thing. These two groups of people think differently. There are different processes going on in their minds. They will reach different results. Even if neither side suddenly breaks out a test tube, one side will be doing something fundamentally more scientific than the other.

And let me show my colors: I think one of them is doing something *better*. I myself am a scientismist. I think the

impact of having people thinking scientifically in non-scientific fields is usually good.

II.

Why should that be? If science is just about rocks and molecules and stars, why would scientific training and knowledge give you an advantage in unrelated fields?

From Shakespeare's Cassius: "The fault, dear Brutus, is not in our stars / But in ourselves."

I don't think science should inform philosophy because of what it's discovered about stars. It should inform philosophy because of what people in the process of investigating stars have incidentally discovered about the faults in themselves.

Imagine a prankster with superhuman skill in ophthamological surgery manages to cut open and rearrange your eyes while you're asleep. She gives your vision a sort of tilt-shift effect that makes everything appear smaller. And at the time, you happen to be on a World Tour.

Your friend asks you how Paris is, and you say: "It looks very small! It's full of tiny people and a miniature Eiffel Tower!" Your friend corrects you and tells you Paris is actually normal sized.

Then you're in London. You mention how it's full of dwarves and a cute little clock tower the size of a sewing needle. Once again your friend corrects you and tells you London is normal size.

The next week you're in Beijing. You're tempted to dismiss it as a city of midgets and of medium-sized portraits of Mao. But by now you've wised up. Your experiences in Paris and London have taught you that there's something wrong with your vision and you had better be more careful.

A detractor might say “What can learning about Paris and London possibly teach you about Beijing? It’s on a totally different continent and steeped in a totally different culture. Lessons learned in Europe just don’t transfer!” But as long as you’re using the same faulty vision to view each city, the lessons learned *do* transfer. Even if facts about China are completely uncorrelated with any facts about Europe, your *errors* about both will be correlated because it’s the same person erring each time.

III.

For all that it stresses empiricism, science isn’t just about experiment and observation. It’s also got a theoretical side. The interesting part of science is that it’s a calibration process. You use your theorizing faculties, and then you perform experiments to see if you were right or wrong.

Just as a biologist-engaged-in-experiment is testing different drugs to see whether they cure disease, a biologist-engaged-in-theory is (usually unintentionally) testing different mental algorithms to see whether they correctly predict which drugs will cure disease, or can generate disease cures.

And just as experimental science may discover that the witch doctor’s technique of drilling a hole in the skull to let out the evil demons is not in fact best practice, so theoretical science may discover that certain reasoning techniques don’t stand up to scrutiny either.

One of these which has been downright mythologized is the story of How We Learned That Things Aren’t Usually Caused By Sentient Agents. Back in the old days rain was caused by the Rain God and disease was caused by the Disease Demon, but then we discovered that these were actually natural processes and not people at all, and (so the myth continues)

One Day We Will Finally Complete The Process By Ceasing To Believe In God.

The only problem with this narrative is that as far as I know we stopped believing in the Rain God and the Disease Demon long before we had any good experimental science or even any naturalistic alternative explanations for rain or disease. I'm not sure why this is, but it makes it less than a perfect victory for Science.

Still, some very similar stories *are*. The Copernican Principle, for one, where we gradually lost belief in our own uniqueness and went from "Earth holds a privileged position" to "The solar system holds a privileged position" to "The galaxy holds a privileged position" to our current and obviously-correct "Okay, there's a whole universe out there, but it *definitely* has a privileged position and doesn't split up into lots of different equally real quantum branches".

There are other principles without equally catchy names. The "No, You Can't Just Treat Human-Level Interesting Categories As Ontologically Real Primitives" principle, which I suppose one could call the Huxleyan Principle after the biologist who worked the hardest to discredit elan vital. The "Stop Using Value-Based Explanations" principle, which can be used with equal aplomb against everyone from the old Great Chain of Being theorists to high school biology students who insist that evolution is progress from "worse" to "better" organisms.

(life hack: Does saying "worse" and "better" make you feel unscientific? Just replace these words with "less complex" and "more complex", then pretend these terms have objective meanings!)

IV.

Each of these principles works not because of the particular field it is applied to, but because it compensates for a defect in our own reasoning faculty. Our brain evolved mostly to think about other humans, thinking about other humans is the first thing it wants to do in any situation, so we end up with a bias towards anthropomorphism. We are clearly very important to ourselves, so we project this onto the universe and think we (or our planet, or our star, or our galaxy) must be at the center.

Therefore, the correct application of these principles is an antiprediction, a sort of easily defensible sticking to a default position. For example, "the world will probably not end on January 18, 2020" is an antiprediction, because we have no reason to think that it should. It is very difficult to predict the future, but this is no argument against my claim that the world will probably not end on January 15, 2020. No one gets to shake their head and say "That's kind of *arrogant* of you to think that you can know that."

Antipredictions do not always *sound* like antipredictions. Consider the claim "once we start traveling the stars, I am 99% sure that the first alien civilization we meet will *not* be our technological equals". This sounds rather bold – how should I know to two decimal places about aliens, never having met any?

But human civilization has existed for 10,000 years, and may go on for much longer. If "technological equals" are people within about 50 years of our tech level either way, then all I'm claiming is that out of 10,000 years of alien civilization, we won't hit the 100 where they are about equivalent to us. 99% is the exact right probability to use there, so this is an antiprediction and requires no special knowledge about aliens to make.

The antipredictive nature is surprising because certain possibilities stand out more clearly to us. Aliens being around our tech level is *narratively interesting* – we can fight wars on an equal footing or engage in mutually profitable trade. Certainly the idea of meeting aliens who have been stuck at the tech level of Assyria for a thousand years is less available.

We can say that the hypothesis-space is distorted: that equal-tech aliens *looks* like it takes up a very large area, even though it is tiny.

In the same way, certain salient regions of hypothesis space that correspond to natural human thought processes falsely appear very large, and certain other regions that don't correspond to natural thought processes falsely appear much smaller.

If you've calibrated yourself on previous problems, then "The ground of being has to be a person" should bring up alerts like "Wait a second, it also seemed like rain had to be a person".

And "I bet moral value is this objectively real conceptual primitive of perfect simplicity" should bring up alerts like "Wait a second, it also seemed like life had to be an objectively real conceptual primitive of perfect simplicity, and it ended up being this."

This should work the same way that the observation "Beijing seems full of tiny little Chinese midgets" brought up the alert "Wait a second, Paris also seemed full of tiny little French midgets".

V.

Beyond these specific problems like the Copernican Principle lies a greater a problem which makes all others pale into insignficance.

People who haven't calibrated their theorizing against hard reality still think verbal reasoning works.

There have been a couple hundred proofs of the existence of God thought up throughout the centuries. And more recently, there have also been a couple hundred proofs of the nonexistence of God thought up. Clearly, a couple hundred proofs of something doesn't make it so.

"But no one ever said something must be true just because someone has published a proof! The proof must be correct! The proofs of the existence/nonexistence of God are just wrong!"

Well, yes. Of course. But which side's proofs you think are wrong tend to have a very very very strong correlation with which side you personally subscribe to.

Our faculty for evaluating chains of deductive reasoning similar to proofs of the (non)existence of God, or a lot of what goes on in philosophy, or god help us politics, is – pardon my language – *really shitty*. And we never realize this, because it is selectively shitty. It tells us it has logically evaluated arguments, and determined our opponents' arguments are wrong, and our own arguments are right. And this is nice and consistent and convenient so we assume it must know what it's doing. If it gets proven wrong once or twice or sixty times, we can dismiss that as a fluke, or an edge case, or It's Beside The Point, or The Real Question Is Whether You Are Racist For Even Bringing That Up.

The thing I notice about scientists who branch out into other fields and get accused of scientism is that they tend to be *minimalists*. They're always the ones saying there *isn't* something. There probably isn't a god. There probably isn't Cosmic Consciousness. There probably isn't any particular

moral law beyond your actions just having effects in the world.

And their opponents believe this is because they fetishize Science as the only thing that can possibly be real. You can see bacteria under a microscope, you can see atoms under a microscope, but you can't see God under a microscope, and therefore if they don't believe in God it's because they have obstinately decided only to believe in Science-y things.

But in fact, it is exactly the reverse. These skilled wielders of rejection first trained themselves on Science-y things. Lamarckian evolution. Steady State theory. The planet Vulcan. The four humors. The blank slate. Radical behaviorism. Catastrophism. Recapitulation theory. The luminiferous aether.

By holding scientific theories, which can be and are disproven, they trained themselves in Doubt. And that Doubt continues to serve them when they branch into other areas where theories cannot be disproven so easily. And maybe they will be less easily swayed by attractive verbal arguments.

The people who get accused of scientism are not all themselves scientists, and even those who are may never have suffered a mistake equal in enormity to believing in luminferous aether. But they're steeped in the culture. They've absorbed the mores. Even if they have no scientific virtue themselves are merely aping the motions of their betters, those motions themselves contain certain safeguards against some of the most atrocious errors.

I don't believe such scientifically informed people, when branching off into other fields, will always or even often be right. But I think they have a better chance than people working from intellectual traditions that have never gotten to calibrate their thought processes in the same way.

And that is why I consider myself a scientismist. I know it is supposed to be a perjorative, but I am reclaiming it. And I know it has many definitions, but this one is mine:

A view of hypothesis-space that accounts for human fallibilities, as revealed by past experiences.

And a very, very high burden of proof before zeroing in on any one area of that space.

Whose Utilitarianism?

[Trigger warning: attempt to ground morality]

God help me, I'm starting to have doubts about utilitarianism.

Whose Superstructure?

The first doubt is something like this. Utilitarianism requires a complicated superstructure – a set of meta-rules about how to determine utilitarian rules. You need to figure out which of people's many conflicting types of desires are their true "preferences", make some rules on how we're going to aggregate utilities, come up with tricks to avoid the Repugnant Conclusion and Pascal's Mugging, et cetera.

I have never been too bothered by this in a *practical* sense. I agree there's probably no perfect Platonic way to derive this superstructure from first principles, but we can come up with hacks for it that come up with good results. That is, given enough mathematical ingenuity, I could probably come up with a utilitarian superstructure that exactly satisfied my moral intuitions.

And if that's what I want, great. But part of the promise of utilitarianism was that it was going to give me something more objective than just my moral intuitions. Don't get me wrong; formalizing and consistency-ifying my moral intuitions would still be pretty cool. But that seems like a much less ambitious project. It is also a very personal project; other people's moral intuitions may differ and this offers no means of judging the dispute.

Whose Preferences?

Suppose you go into cryosleep and wake up in the far future. The humans of this future spend all their time wireheading. And because for a while they felt sort of unsatisfied with wireheading, they took a break from their drug-induced stupors to genetically engineer all desires beyond wireheading out of themselves. They have neither the inclination nor even the ability to appreciate art, science, poetry, nature, love, etc. In fact, they have a second-order desire in favor of continuing to wirehead rather than having to deal with all of those things.

You happen to be a brilliant scientist, much smarter than all the drugged-up zombies around you. You can use your genius for one of two ends. First, you can build a better wireheading machine that increases the current run through people's pleasure centers. Or you can come up with a form of reverse genetic engineering that makes people stop their wireheading and appreciate art, science, poetry, nature, love, etc again.

Utilitarianism says very strongly that the correct answer is the first one. My moral intuitions say very strongly that the correct answer is the second one. Once again, I notice that I don't really care what utilitarianism says when it goes against my moral intuitions.

In fact, the entire power of utilitarianism seems to be that I like other people being happy and getting what they want. This allows me to pretend that my moral system is "do what makes other people happy and gives them what they want" even though it is actually "do what I like". As soon as we come up with a situation where I no longer like other people getting what they want, utilitarianism no longer seems very attractive.

Whose Consequentialism?

It seems to boil down to something like this: I am only willing to accept utilitarianism when it matches my moral intuitions,

or when I can hack it to conform to my moral intuitions. It usually does a good job of this, but sometimes it doesn't, in which case I go with my moral intuitions over utilitarianism. This both means utilitarianism can't *ground* my moral intuitions, and it means that if I'm honest I might as well just admit I'm following my own moral intuitions. Since I'm not claiming my moral intuitions are intuitions *about* anything, I am basically just following my own desires. What looked like it was a universal consequentialism is basically just *my* consequentialism with the agreement of the rest of the universe assumed.

Another way to put this is to say I am following a consequentialist maxim of "Maximize the world's resemblance to W", where W is the particular state of the world I think is best and most desirable.

This formulation makes "follow your own desires" actually not quite as bad as it sounds. Because I have a desire for reflective equilibrium, I can at least be smart about it. Instead of doing what I first-level-want, like spending money on a shiny new car for myself, I can say "What I seem to really want is other people being happy" and then go investigate efficient charity. This means I'm not quite emotivist and I can still (for example) be wrong about what I want or engage in moral argumentation.

And it manages to (very technically) escape the charge of moral relativism too. I think of a relativist as saying "Well, I like a world of freedom and prosperity for all, but Hitler likes a world of genocide and hatred, and that's okay too, so he can do that in Germany and I'll do my thing over here." But in fact if I'm trying to maximize the world's resemblance to my desired world-state, I can say "Yeah, that's a world without

Hitler" and declare myself better than him, and try to fight him.

But what it's obviously missing is objectivity. From an outside observer's perspective, Hitler and I are following the same maxim and there's no way she can pronounce one of us better than the other without having some desires herself. This is obviously a really undesirable feature in a moral system.

Whose Objectivity?

I've started reading proofs of an objective binding morality about the same way I read diagrams of perpetual motion machines: not with an attitude of "I wonder if this will work or not" but with one of "it will be a fun intellectual exercise to spot the mistake here". So far I have yet to fail. But if there's no objective binding morality, then the sort of intuitionism above is a good description of what moral actors are doing.

Can we cover it with any kind of veneer of objectivity more compelling than this? I think the answer is going to be "no", but let's at least try.

One idea is a *post hoc* consequentialism. Instead of taking everyone's desires about everything, adding them up, and turning that into a belief about the state of the world, we take everyone's desires about states of the world, then add all of those up. If you want the pie and I want the pie, we both get half of the pie, and we don't feel a need to create an arbitrary number of people and give them each a tiny slice of the pie for complicated mathematical reasons.

This would "solve" the Repugnant Conclusion and Pascal's Mugging, and at least *change the nature* of the problems around "preference" and "aggregation". But it wouldn't get rid of the main problem.

The other idea is a sort of morals as Platonic politics. Hobbes has this thing where we start in a state of nature, and then everybody signs a social contract to create a State because everyone benefits from the State's existence. But because coordination is hard, the State is likely to be something simple like a monarchy or democracy, and the State might not necessarily do what any of the signatories to the contract want. And also no one actually signs the contract, they just sort of pretend that they did.

Suppose that Alice and Bob both have exactly the same moral intuitions/desires, except that they both want a certain pie. Every time the pie appears, they fight over it. If the fights are sufficiently bloody, and their preference for personal safety outweighs their preference for pie, it probably wouldn't take too long for them to sign a contract agreeing to split the pie 50-50 (if one of them was a better fighter, the split might be different, but in the abstract let's say 50-50).

Now suppose Alice is very pro-choice and slightly anti-religion, and Bob is slightly pro-life and very pro-religion. With rudimentary intuitionist morality, Alice goes around building abortion clinics and Bob burns them down, and Bob goes around building churches and Alice burns them down. If they can both trust each other, it probably won't take long before they sign a contract where Alice agrees not to burn down any churches if Bob agrees not to burn down any abortion clinics.

Now abstract this to a civilization of a billion people, who happen to be divided into two equal (and well-mixed) groups, Alicians and Bobbites. These groups have no leadership, and no coordination, and they're not made up of lawyers who can create ironclad contracts without any loopholes at all. If they had to *actually* come up with a contract (in this case maybe

more of a treaty) they would fail miserably. But if they all had this internal drive that they should imagine the contract that would be signed among them if they could coordinate perfectly and come up with a perfect loophole-free contract, and then follow that, they would do pretty well.

Because most people's intuitive morality is basically utilitarian [citation needed], most of these Platonic contracts will contain a term for people being equal even if everyone does not have an equal position in the contract. That is, even if 60% of the Alicians have guns but only 40% of the Bobbites do, if enough members of both sides believe that respecting people's preferences is important, the contract won't give the Alicians more concessions on that basis alone (that is, we're imagining the contract real hypothetical people would sign, not the contract hypothetical hypothetical people from Economicsland who are utterly selfish would sign).

Whose Communion?

So what about the wireheading example from before?

Jennifer RM has been studying ecclesiology lately, which seems like an odd thing for an agnostic to study. I took a brief look at it just to see how crazy she was, and one of the things that stuck with me was the concept of communion. It seems (and I know no ecclesiology, so correct me if I'm wrong) motivated by a desire to balance a desire to unite as many people as possible under a certain banner, with the conflicting desire to have everyone united under the banner believe mostly the same things and not be at one another's throats. So you say "This range of beliefs is acceptable and still in communion with us, but if you go outside that range, you're out of our church."

Moral contractualism offers a similar solution. The Alicians and Bobbites would sign a contract because the advantages of coordination are greater than the disadvantages of conflict. But there are certain cases in which you would sign a much weaker contract, maybe one to just not kill each other. And there are other cases still when you would just *never sign a contract*. My Platonic contract with the wireheaders is "no contract". Given the difference in our moral beliefs, whatever advantages I can gain by cooperating with them about morality are outweighed by the fact that I want to destroy their entire society and rebuild it in my own image.

I think it's possible that all of humanity except psychopaths are in some form of weak moral communion with each other, at least of the "I won't kill you if you don't kill me" variety. I think certain other groups, maybe along the culture level (where culture = "the West", "the Middle East", "Christendom") may be in some stronger form of moral communion with each other.

(note that "not in moral communion with" does not mean "have no obligations toward". It may be that my moral communion with other Westerners contains an injunction not to oppress non-Westerners. It's just that when adjusting my personal intuitive morality toward a morality I intend to actually practice, I only acausally adjust to those people whom I agree with enough already that the gain of having them acausally adjust toward me is greater than the cost of having me acausally adjust to them.)

In this system, an outside observer might be able to make a few more observations about the me-Hitler dispute. She might notice Hitler or his followers were in violation of Platonic contracts it woud have been in their own interests to sign. Or

she might notice that the moral communions of humanity split neatly into two groups: Nazis and everybody else.

I'm pretty sure that I am rehashing territory covered by other people; contractualism seems to be a thing, and a lot of people I've talked to have tried to ground morality in timeless something-or-other.

Still, this appeals to me as an attempt to ground morality which successfully replaces obvious logical errors with complete outlandish incomputability. That seems like maybe a step forward, or something?

EDIT: Clarification in my response to Kaj here.

Book Review: After Virtue

A few weeks ago the blogosphere discovered Ayn Rand's margin notes on a C.S. Lewis book. They were everything I expected and more. Lewis would make an argument, and then Rand would write a stream of invective in the margin about how much she hated Lewis' arguments and him personally. I kind of wanted to pat her on the shoulder and say "Look, I'm really sorry, but *he can't hear you*."

But I can also sympathize with her. It is *infuriating* to read a book making one horrible argument after the other. And when it glibly concludes "…and therefore I am right about everything", and you know you'll never be able to contact the author, it gives a pale ghost of satisfaction to at least scrawl in the margin "YOUR ARGUMENTS ARE BAD AND YOU SHOULD FEEL BAD".

This is kind of how I felt about Alasdair MacIntyre's *After Virtue*.

As far as I can tell, MacIntyre's central argument works something like this:

1. There are many theories of ethics in existence today
2. The ones that came after Aristotelianism have failed to objectively ground themselves and create a perfect society in which everyone agrees on a foundation for morality
4. Therefore, we should return to Aristotelianism

You may notice a hole where one might place a Step 3, something like "Aristotelianism, in contrast, *did* objectively ground itself and create a perfect society in which everyone agreed on a foundation for morality." This is exactly the argument MacIntyre digresses into a lengthy explanation of

how much he likes Greek tragedy to hope we will avoid noticing him not making.

To MacIntyre's credit, he does a pretty good critique of modern moral philosophy. He says that since society doesn't share any kind of moral tradition, we can debate important moral questions – like abortion, or redistributive taxation – until the cows come home, but this is in fact only the appearance of debate since we have no agreed-upon standards against which to judge these things. Because we cannot settle these by rational argument, instead we turn to outrage and attempts to shame our opponents, making the protester one of the archetypal figures of the modern world.

("…making the [unsavory sounding figure] one of the archetypal figures of the modern world" is one of MacIntyre's pet phrases. It starts grating after a while.)

I broadly agree with him about this problem. I discuss it pretty explicitly in sections 6.5 and 8.1 of my Consequentialism FAQ. I propose as the solution some form of utilitarianism, the only moral theory in which everything is commensurable and so there exists a single determinable standard for deciding among different moral claims.

Annnnnd MacIntyre decides to go with virtue ethics.

The interesting thing about virtue ethics is that it is *uniquely bad at this problem*. In the entire book, MacIntyre doesn't give a single example of virtue ethics being used to solve a moral dilemma, as indeed it cannot be. You can attach a virtue (or several virtues) of either side of practically any moral dilemma, and virtue ethics says exactly nothing about how to balance out those conflicting duties. For example, in Kant's famous "an axe murderer asks you where his intended victim is" case, the virtue of truthfulness conflicts with the virtue of

of compassion (note, by the way, that no one has an authoritative list of the virtues and they cannot be derived from first principles, so anyone is welcome to call anything a virtue and most people do).

MacIntyre totally admits this conflict, but instead of saying it's a problem with his theory he says it's the tragedy of human existence, then says that the virtue of justice is knowing how to balance those two virtues.

So basically, his entire condemnation of all systems beside his own is based on the difficulty of coming to moral consensus, but his own means of coming to moral consensus is a giant black box labelled "THE VIRTUE OF BEING ABLE TO SOLVE THIS HERE PROBLEM CORRECTLY".

I don't like deontology. In fact, I dislike it more than almost anyone I know except maybe Federico. But I will give credit where credit is due: deontology actually comes up with solutions to moral problems. The solutions are wildly incorrect and incredibly harmful, but they get a gold star for effort.

Virtue ethics, as far as I can tell, just gives you a knowing look and says "The very fact that you interpret morality in terms of *moral dilemmas* is a symptom of the disease of liberal modernity." This is useful for sounding deeply wise, but little else. If you ask "Okay, but disputes over morality are an actual feature of the real world, and the whole reason we're doing this ethics stuff is to try to solve them, so if we admit we're diseased and the ancient Greeks were awesome, maybe you could help us out here?" – then virtue ethics just takes another sip of wine from its table in the corner and says "Your decadent individualist mind has no idea how disappointed Aristotle would be in you for even asking that. Did you even *consider* just being a virtuous city-state in which everyone is a

great-minded soul acting for the good of the polis? I didn't *think* so."

If You Can't Convince 'Em, Just Start Reciting The Entire History Of The Human Race

Beyond my distaste for *After Virtue*'s philosophy, I wasn't a huge fan of its history either.

The book claims that the reason we don't have a working agreed-upon morality is that the ancient Greeks (and medievals) *did* have a working agreed-upon morality (virtue ethics), but when it collapsed we were left with all these weird phrases like "virtuous" and "should" and "ought" and "the good" and outside the context of virtue ethics had no idea what to do with them. Since we couldn't use the correct virtue-ethics solution, we entered the age of interminably debating what the correct solution was, hence the modern age of moral dilemmas.

In fact, the beginning of the book is a fascinating and attractive metaphor (drawn from the excellent *A Canticle For Leibowitz*) in which all scientific knowledge is destroyed by some apocalypse. A future civilization picking over the scraps forms a sort of cargo cult in which they know there are supposed to be things called "electrons", and that the equation "e = mc^2" is very important for no reason, but no matter how many times they debate what shape these "electrons" were supposed to be or whether the c in e=mc^2 stands for 'color' or 'correctness', they can't seem to produce rockets or nuclear power. Phrases like e=mc^2 only make sense as part of a tradition; a stupid debate about whether c stands for color or correctness is a symbol that we're trying to interpret it separately from that tradition and we're just going to end up confusing ourselves. To MacIntyre, the tradition here is virtue

ethics and modern society plays the role of the postapocalyptics looking quizzically over the scraps.

(the apocalypse? The Enlightenment, of course. Just *once* I want to go a whole week without someone blaming everything on the Enlightenment.)

Alasdair MacIntyre is clearly an expert classical scholar. And in fact he discusses the classical world's disputes on morality very competently in his book. So it bewilders me that he doesn't notice that actually, modern society's debates over the Good are no different than those of the classical world. He even cites Sophocles' tragedy Philoctetes as an example of moral dilemma in the ancient world. I agree – it is a perfect moral dilemma – of exactly the sort MacIntyre is claiming only exists because our civilization is living in the postapocalyptic ruins of virtue ethics. And *Philoctetes* was written twenty years before Aristotle was even born. Heck, forget Sophocles, even Socrates is a perfect example of this kind of moral inquiry.

MacIntyre then waxes about the wonder of the Greek city-states, which he says were communities where everyone was united on a single view of the good – that which was the proper *telos* of man.

Except, once again, all the problems of the modern age appear in the Greek city-states as well. Athens went from the laws of Solon to the tyranny of Peisistratus to the dictatorship of Hippias to the democracy of Cleisthenes to the oligarchy of the Four Hundred to the Thirty Tyrants to the democracy of Thrasybulus all in about a century. The periods of democracy were as rife with hostile factions and unresolved issues as any period in modern America or Europe.

The idea that everyone back then was happily united around the Objectively Proper End of Man is slightly complicated by the fact that no one back then agreed on what the Objectively Proper End of Man was, any more than anyone today agrees on what the Proper End of Man is, least of all virtue ethicists and super-dog-double-least of all anyone who reads the book *After Virtue* which happily informs us that pursuing it will solve all our problems but neglects to mention what the heck it might be or give us a shred of evidence to overcome our high priors against such a thing existing.

Then there's a short focus on the medieval period, which I am told is marked by everyone being very virtuous but otherwise not particularly worthy of remark, followed by an attack on David Hume and Immanuel Kant, who apparently both *totally failed to be virtue ethicists*.

The modern period is marked…okay, I understood this part even less than the other parts. The modern period is marked by the Bureaucrat, who is another one of those Archetypal Figures Of The Modern World (others include the Aesthete and the Therapist). The Bureaucrat claims to have expertise in some subject, but clearly this is a lie, because no one can ever understand human affairs infallibly and this is *kind of* like saying no one can ever understand human affairs at all. Since everyone loves bureaucrats, who are people who claim to be able to understand human affairs, and yet no one can *really* understand human affairs, something must be wrong, and for all we know that something could be that we're not all virtue ethicists (am I strawmanning here? Read pages 79-108 and find out).

Somebody Here Is Really Confused, And I Just Hope It's Not Me

I have never been able to appreciate Continental philosophy (well, Nietzsche was pretty cool, but I have a hard time classifying anyone who can actually write engagingly as a Continental philosopher). *After Virtue*, despite having been written by a verified Scotsman by all accounts closely engaged with the analytic tradition, just seemed really Continental to me. It avoided logical arguments for a particular well-defined point in favor of long historical meanderings carefully designed to make the reader vaguely worry that everything was socially constructed and that the reader's social construction was particularly rotten, without ever coming out and explicitly saying anything that could be seized upon as a claim to evaluate.

But the thing is that MacIntyre is considered one of the greatest living philosophers, and *After Virtue* one of the century's greatest works on ethics. Just on priors I'm more likely to be misunderstanding him than he is to be talking nonsense. Even people I respect – including Catholics from the Patheos community and a few rationalists from the Less Wrong community – recommend MacIntyre.

Those same people recommended Edward Feser to me. There are a lot of similarities between Feser and MacIntyre – both say that the philosophical tradition of Greece and the medieval age was much better than our own tradition, and that we're so screwed up we can't even *realize* how screwed up we were. Both have very good things to say about teleology, and both ended up Catholic as a result of their philosophical studies.

I really enjoyed Feser's *The Last Superstition* (and his *Aquinas*, although that's less relevant here). I thought it did a great job bridging a wide inferential gap and really illuminated why he thought the things he thought. I think his account of forms and teleology is flawed because of a few basic errors in

his foundations (I started explaining why on my old blog but never really finished) but it was flawed in ways where I could understand the force of his arguments and why his premises would lead to that conclusion. Even if I ended up disagreeing with his answers, I gained a huge admiration for his ability to ask the right questions and go about investigating them in the right way.

But as his occasional enemy Chris Hallquist delights to point out, Feser is not a hugely prestigious figure in mainstream academic philosophy. MacIntyre is. I was hoping for the same fascinating ideas, but with a suave British cool instead of hilarious over-the-top rants. Instead I got…I don't even know.

I am really sorry, virtue ethicists. But you are going to have to do better than this if you want me to understand you.

Read History of Philosophy Backwards

I disagreed with the specific presentation of history of philosophy in *After Virtue*, but not with its decision to present history of philosophy.

This is new for me. Back when I was in college, my chief complaint about my philosophy course was that it spent all its time teaching things that Aristotle or Plato or Descartes thought that were just *obviously wrong*. I sort of annoyed my professors by constantly raising my hand and being like "Sorry, isn't Plato completely confused here because now we know that actually X", and my professor would be "Well, that's a very interesting theory" and then continue teaching Plato.

None of these classes were billed as "history of philosophy", but as "philosophy" itself. I knew better than to expect to be taught a single thing that was definitely right, but I had kind of hoped they would limit it to things that had some chance of being true, or that couldn't be seen through by a bright undergraduate.

As Dave Barry puts it:

> "I was terrible at history. I could never see the point of learning what people thought back when people were a lot stupider. For instance, the ancient Phoenicians believed that the sun was carried across the sky on the back of an enormous snake. So what? So they were idiots"

I still believe that if you only have four years to teach an undergraduate philosophy, there's no way you should be

teaching this kind of stuff before you teach them genuinely useful things like what the heck concepts are and why you can't suddenly change the moral value of things by calling them different names. But I no longer think this is quite as useless as I previously believed.

Today I was discussing Sartre with a friend, and a lot of the discussion centered around why people care about Sartre. Sartre's main point – that no one else can tell you who you are, and you choose what your own values are – seems so cliched, so much like what an uncreative graduation speaker might say – that it hardly seems worth elevating him to the Canon Of Philosophical Greatness.

My hypothesis – and I don't know if it's true – is that this is only cliched now because Sartre won. The point of studying Sartre is not to learn that you choose your own identity, but to *read him backward* – to start with this idea that choosing your own identity is obvious, and then read Sartre to learn exactly how controversial it was at the time and what sorts of arguments Sartre had to go through to get people to accept it, and eventually understand the position that the original reader of Sartre was supposed to have *started with*. If you succeed, you might still believe that you choose your own identity, but you'll also understand that this isn't an obvious necessary fact of the universe, that there used to be people who believed you didn't and that they had some good arguments too.

Sometimes this is true even when you don't know that it's true. When I first studied Hobbes in college, I was under the impression that nobody agreed with Hobbes these days – after all, Hobbes was a believer in absolute monarchy, and now everyone is strongly opposed to that. But later I realized that pretty much everyone is a Hobbesian in that Hobbes was one of the first people to think in terms of people coming together

to found a government for their mutual self-interest; previously governments were either just the natural state of human affairs, or part of the hierarchical nature of the universe under God, or composed because the *telos* of man only flourishes in a community, or not even something you thought about. Indeed, Alasdair MacIntyre seems to be at least partially advocating a return to pre-Hobbesian ideas about government, even though he doesn't put it in those terms.

The only reason I had Hobbes pegged as "the absolutism guy" was because that was the only place in which his theories differed from my own and so I assumed it was his only idea that wasn't "obvious". If I had read him backward, I would have gotten a lot more out of him.

Under this model, reading philosophers who were completely wrong is another way of unlearning your assumptions. For example, I originally thought the term "reductionism" was essentially meaningless; the opposite of "reductionism" was "not thinking things through clearly and having incoherent ideas". After I read Aristotle, I changed my mind; he proposes a non-reductionism which for all I know very well may be the case in Dimension Q'qaar, even though it has no relation to how the real world works.

Under this model, the point of reading history of philosophy is to unlearn your assumptions. Growing up in a certain cultural tradition not only infuences the answers you think are right, but the potential answers you're able to generate and even the questions you're able to ask.

For some people this is important because the past was actually correct or close to it. I hang out with a disproportionate number of these people. I was briefly taken aback when Chris claimed here that he *doesn't* have to

constantly listen to claims that the Enlightenment ruined everything.

But what if, like me, you think the past was pretty thoroughly wrong about most philosophical issues?

Then I *still* think it's important to have a non-parochial worldview because the *next* big idea is likely to be just as different from present philosophy as present philosophy is from past philosophy. And unless you realize how different present philosophy is from past philosophy, you won't even have the mental mechanism to expand your search space large enough to capture something worthy of participating in the future.

Virtue Ethics: Not Practically Useful Either

I've been trying to understand some of the responses to my review of *After Virtue*. Tell me whether this is about right:

The problem of "doing the right thing" consists of two subproblems.

First, knowing what the right thing is. Do we legalize or ban abortion? Do we press the switch in the trolley problem or not?

Second, behaving correctly in situations where we do know what the right thing is. For example, going to visit a friend in the hospital even though the hospital is far away. Not cheating on your taxes even though you could use the money. Working hard even though no one is checking up on you.

The proposal was that virtue ethics doesn't claim to be a solution to the first problem, but is a uniquely excellent solution to the second, and in fact the solution people actually use. Am I understanding this correctly?

Because if that's true, I *still* disagree.

Virtue Ethics Is What People Do

I am not really good at thinking in terms of good or bad people. I can do so in very edge cases, like Kim Jong-il or St. Holden. But I tend to process all the people I know and have remotely okay interactions with as "good people", leading to conversations like:

Me: Oh, cool, Bob is coming over soon! Bob is great!
Friend: But didn't Bob [do X, Y, and Z]?
Me: Well, yes, but other than that he's great.

Friend: And didn't he [do A, B, and C]?
Me: You can't just keep taking all these things Bob does out of context!
Friend: Okay, what has Bob done that was good?
Me: He…well…he…you know! Bob! He's great!

When I deviate from this, it almost always tends to be in terms of actions, not qualities. Like "Bob is great, except that he posts really annoying things on Facebook all the time". Or "Bob is great, except that he has some really horrible politics."

When I think about my own morality, it's almost never in terms of whether I am or am not a virtuous person (I have at least one subagent that's always convinced I'm a virtuous person, and at least one subagent that's always convinced I'm terrible and deserve to die, and doing good things paradoxically strengthens the latter subagent for some reason).

It's usually in terms of – I guess it feels like a missile must feel locking on to a target. If my mind is sufficiently calm and predisposed to goodness, I think "Wait a second, *that's* the morally correct thing to do", lock on to one option, and then feel really good about it – a serene feeling.

I can't always do this. If I'm sufficiently angry, part of me thinks "I bet I could lock on to what's good and do it…but that would probably involve turning the other cheek, or compromising, AND THEN THESE JERKS WOULD GET AWAY WITH IT." And then my mind makes up all sorts of game theoretic justifications for why it's more important to punish defectors than to do what feels like the right thing at this precise moment. Or if I'm sufficiently exhausted, I think "If I started worrying about what's good now, then I'd probably have to do it, and that would be really arduous."

On the one hand, this is no doubt a very idiosyncratic report; I don't expect my experience of morality to be similar to anyone else's. On the other hand, this is an idiosyncratic report and I don't expect my experience of morality to be similar to anyone else's. So if you say "Virtue ethics is the way people naturally think about morality!", that's either a typical mind fallacy or you're going to have to do a *much* better job explaining virtue ethics.

My experience of morality is contray to traditional virtue ethics in almost every way. It doesn't feel like it depends on my social roles. It doesn't strike me as divisible – that is, it feels like solid goodness and words like "continence" or "prudence" don't do anything to me. It doesn't strike me as the same feeling that occurs when I consider important but non-ethical questions like procrastination. It doesn't strike me as performed in a community or according to a narrative. It's just *not* virtue ethics.

The Practice of Making People More Moral

The other claim is that virtue ethics is the science of making people better – a process for helping people refine their existing moral intuitions and overcome temptation more effectively.

If that's true, it's a science much like medieval medicine was a science – totally untested and not especially likely to bear any resemblance to reality. If people are *actually* looking for ways to become more moral, I bet an hour's search would find about thirty of them that are more likely to work than adopting virtue ethics.

These could be broadly divided into beliefs and practices. In terms of beliefs, I think the most useful would be a belief in

the Devil, moral realism, humility, and various kinds of magical thinking.

In terms of practices, there would be willpower training (pretty much anything that requires willpower counts as willpower training, but let's say exercise as the stereotypical example), rationality training, keeping a journal, talking about morality, making friends, joining a group with some sort of interest in morality, cutting yourself off from bad influences, making yourself happier (happy people are more moral), learning relaxation/stress-busting techniques, and reading fiction.

All of these things would make people more moral in different directions and in different ways. For example, I bet reading works of fiction about poor people in the Third World would make you more likely to donate to charity, and contemplating virtue ethics and the just *polis* would make you more likely to make you get involved in local politics. Which of these you recommend is very closely linked to whether you think giving to charity to the Third World is more or less important than getting involved in local politics (hint: there is only one answer to this question which is not really stupid).

In terms of the best all-around practice for increasing morality I would nominate meditation, especially lovingkindness meditation. David Chapman, who knows ten zillion times more about Buddhism and meditation than I do, suggests metta bhavana, tonglen, and chöd. Even very generic meditation ticks several of the boxes above – relaxation, willpower training, and happiness – but these are said to (and have some evidence of) specifically increasing your ability to love and care about other people.

This seems like probably the best thing you can do for morality short of ground it objectively. If people love and care

about others, they end up automatically ticking the two boxes I claim are required to end up more-or-less utilitarian – grounding morality in the world and caring about other people. If you've got that, it kind of lowers the degree to which morality even needs to be grounded objectively; it would still be nice, but we can trust people to do the right thing even if it isn't.

Virtue ethics doesn't satisfy either of these criteria, and in fact, we find that throughout history a lot of really terrible people have been very good virtue ethicists (the Spartans come to mind). So although many of the commenters here want to virtue ethics from its failure to ground morality by saying it removes the need to ground morality, I think virtue ethics can't even do that right and there are a lot of things that are much better.

Edit: Something I said in the comments that might clarify my position. I think even this is giving virtue ethics too much credit, since it's *not* just "use our inborn moral sense" but a host of claims about making lists of virtues and studying teleology – but on the principle of steelmanning an opponent's argument:

> Imagine that instead of virtue ethics we're talking about grammar. In most cases, we have a natural grammar sense – that is, the real reason I don't say "Me is Scott" is because it just sounds wrong.
>
> In most cases this is good enough. In some cases it isn't – for example, sometimes we have to teach grammar to foreigners who lack this intuitive grammar sense. Or sometimes there are edge cases where we're really not sure what word to use. Or we want to program a computer to write with proper grammar. Or we want to

set editorial policy for a newspaper. Or sometimes we're just genuinely curious how grammar works.

There is a point to having a science of grammar where smart people say "Oh, it looks like the predicate nominative form is used in this way."

And inventing a "virtue grammar", where people say "But you're ignoring all normal grammar usage in favor of a few silly edge cases! Real people don't talk about predicate nominatives! Just use your natural grammar sense!" is a total waste of everyone's time. Yes, natural grammar sense usually works well, but shouting "Hey, natural grammar sense often works well!" contributes nothing to the field and is just distracting people from actually figuring out how grammar works.

In pretty much all fields except ethics, everyone has agreed that the proper thing to do is to be happy when our natural senses are good enough, but also create a formal study of the field in order to go beyond what our natural senses can tell us. I don't understand why we can't also do this for ethics.

Last Thoughts on Virtue Ethics

The discussion on the other posts has sort of degenerated into people pointing out that our intuitive moral sense is a whole lot more useful most of the time than the speculations of moral philosophers, therefore virtue ethics.

I have two complaints here, the first of which is that virtue ethics is *not* just the claim that we should use our intuitive moral sense. It makes highly counterintuitive or controversial claims like the following:

1. Ethics involves teleology, eg considering the objectively proper ends of beings
2. Ethics has to be grounded in a community to make sense; individual ethics are only a pale shadow
3. Ethics is role-dependent; your role as a mother or child or employee or citizen produces your ethical obligations
4. Ethics is better thought of as about people's character than about the acts they perform
5. It is useful and important to subdivide good behavior into certain virtues like justice, wisdom, and fortitude

1 is almost universally disagreed with by everyone not a practicing virtue ethicist in a philosophy department or a very theologically-minded Catholic. 2 and 3 seem like things most people have no strong opinion about and would leave it for philosophers to debate. You could cherry-pick examples of people's behavior where it looks like they believe 4 and 5 (we have phrases like "bad things happen to good people" which implies we thing in terms of good people) but you could equally well cherry-pick examples of people's behavior where it seems they believe the opposite (the phrase "doing a good deed" implies that we think in terms of good actions).

So none of the five major claims of virtue ethics make it "just our intuitive morality". This is an attempt to load the scales by privileging your own position, like the Muslims who claim that everyone is born a Muslim and it's only when children are brainwashed by their societies that they become anything else.

So if we stop calling it "virtue ethics" and call it a better name like "intuitive ethics", is there any value to the claim "just use your intuitive morality"? Sooooort of, but not the type of value that is actually, well, valuable.

We can use our intuitive morality to determine we should not go around murdering little kids for no reason. This is good. But as a consequence, *no one is remotely interested in the question of whether we should go around murdering little kids for no reason*. No one goes to moral philosophers to ask that question. The very fact that it is solvable by intuitive ethics means that *it is a solved problem*.

The *only* reason anyone is interested in moral philosophy is because sometimes this doesn't work. Maybe we have sociopaths who are mysteriously born without intuitive morality. Or we have controversial moral problems like abortion where people intuitive moralities give very different answers. Or we have difficult moral problems like the Trolley Problem where many people's intuitive moralities just go "Hmmm, that's a really tough question". Or we notice that in olden times, people's intuitive moralities told them slavery was a-ok, including people like Aristotle who had put a lot of work into cultivating their facility of judgment, and we want to make sure we're not doing something equally awful ourselves.

To answer "Use your intuitive morality" in any of these cases ignores the fact that the set of problems where we need moral

advice is the exact complement of the set of problems where using your intuitive morality is good enough.

And the set of senses in which "well, just apply as much intuitive morality as you can and hope it works" solves these kinds of problems is the exact complement of the set of senses in which people still feel like the problem needs to be solved.

If the only claim of "virtue" ethicists is "in the subset of problems where our intuitive morality gives clear and uncontroversial results, great, let's go with those" then I agree with this claim.

If they claim any of statements 1-5 above, or that this generalizes to the case of difficult moral problems or problems anyone actually wants answered, they are going to need to present the evidence that I still maintain *After Virtue* lacked.

Proving Too Much

The fallacy of Proving Too Much is when you challenge an argument because, in addition to proving its intended conclusion, it also proves obviously false conclusions. For example, if someone says "You can't be an atheist, because it's impossible to disprove the existence of God", you can answer "That argument proves too much. If we accept it, we must also accept that you can't disbelieve in Bigfoot, since it's impossible to disprove his existence as well."

I love this tactic *so much*. I only learned it had a name quite recently, but it's been my default style of argument for years. It neatly cuts through complicated issues that might otherwise be totally irresolvable.

Because here is a fundamental principle of the Dark Arts – you don't need an argument that can't be disproven, only an argument that can't be disproven in the amount of time your opponent has available.

In a presidential debate, where your opponent has three minutes, that means all you need to do is come up with an argument whose disproof is inferentially distant enough from your audience that it will take your opponent more than three minutes to explain it, or your audience more than three minutes' worth of mental effort to understand the explanation.

The noncentral fallacy is the easiest way to do this. "Martin Luther King was a criminal!" "Although what you say is technically correct, categories don't work in the way your statement is impl – " "Oh, sorry, time's up."

But pretty much anything that assumes a classical Aristotelian view of concepts/objects is gold here. The same is true of any

deontological rules your audience might be attached to.

I tend to get stuck in the position of having argue against those Dark Artsy tactics pretty often. And the great thing about Proving Too Much is that it can demolish an entire complicated argument based on all sorts of hard-to-tease-apart axioms in a split second. For example, *After Virtue* gave (though it does not endorse) this example of deontological reasoning:

> I cannot will that my mother should have had an abortion when she was pregnant with me, except perhaps if it had been certain that the embryo was dead or gravely damaged. But if I cannot will this in my own case, how can I consistently deny to others the right to life that I claim for myself? I would break the so-called Golden Rule unless I denied that a mother in general has a right to an abortion.

It seemed unfair for me to move on in the book without at least checking whether this argument was correct and I should re-evaluate my pro-choice position. But that would require sorting through all the weird baggage here, like what it means to will something, and whether your obligations to potential people are the same as your obligations to real people, and how to apply the Golden Rule across different levels of potentiality.

Instead I just thought to myself: “Imagine my mother had raped my father, leading to my conception. I cannot will that a policeman had prevented this rape, but I also do not want to enshrine the general principle that policemen in general have no right to prevent rape. Therefore, this argument proves too much.” It took all of five seconds.

Sometimes a quick Proving Too Much can tear apart extremely subtle philosophical arguments that have been debated for centuries. For example, Pascal's Wager also proves Pascal's Mugging (they may both be correct, but bringing the Mugging in at least proves ignoring their correctness to be a reasonable and impossible-to-critique life choice). And Anselm's Ontological Argument seems much less foreboding when you realize it can double as a method for creating jelly donuts on demand.

Interestingly, I think that one of the examples of proving too much on Wikipedia can itself be demolished by a proving too much argument, but I'm not going to say which one it is because I want to see if other people independently come to the same conclusion.

IX. Liberty

The Non-Libertarian FAQ (aka Why I Hate Your Freedom)

Introduction

0.1: Who are you? What is this?

You can find more information about me at www.raikoth.net. This is the second version of the Non-Libertarian FAQ (aka Why I Hate Your Freedom). You can find the original version, which is shorter but more readable, here.

0.2: Are you a statist?

No.

Imagine a hypothetical country split between the "tallists", who think only tall people should have political power, and the "shortists", who believe such power should be reserved for the short.

If we met a tallist, we'd believe she was silly - but not because we favor the shortists instead. We'd oppose the tallists because we think the whole dichotomy is stupid - we should elect people based on qualities like their intelligence and leadership and morality. Knowing someone's height isn't enough to determine whether they'd be a good leader or not.

Declaring any non-libertarian to be a statist is as silly as declaring any non-tallist to be a shortist. Just as we can judge leaders on their merits and not on their height, so people can judge policies on their merits and not just on whether they increase or decrease the size of the state.

There are some people who legitimately believe that a policy's effect on the size of the state is so closely linked to its effectiveness that these two things are not worth distinguishing, and so one can be certain of a policy's greater effectiveness merely because it seems more libertarian and less statist than the alternative. Most of the rest

of this FAQ will be an attempt to disprove this idea and assert that no, you really do have to judge the individual policy on its merits.

0.3: Do you hate libertarianism?

No.

To many people, libertarianism is a reaction against an over-regulated society, and an attempt to spread the word that some seemingly intractable problems can be solved by a hands-off approach. Many libertarians have made excellent arguments for why certain libertarian policies are the best options, and I agree with many of them. I think *this kind* of libertarianism is a valuable strain of political thought that deserves more attention, and I have no quarrel whatsoever with it and find myself leaning more and more in that direction myself.

However, there's a certain more aggressive, very American strain of libertarianism with which I do have a quarrel. This is the strain which, rather than analyzing specific policies and often deciding a more laissez-faire approach is best, starts with the tenet that government can do no right and private industry can do no wrong and uses this faith *in place of* more careful analysis. This faction is not averse to discussing politics, but tends to trot out the same few arguments about why less regulation *has* to be better. I wish I could blame this all on Ayn Rand, but a lot of it seems to come from people who have never heard of her. I suppose I could just add it to the bottom of the list of things I blame Reagan for.

To the first type of libertarian, I apologize for writing a FAQ attacking a caricature of your philosophy, but unfortunately that caricature is alive and well and posting smug slogans on Facebook.

0.4: Will this FAQ prove that government intervention always works better than the free market?

No, of course not.

Actually, in most cases, you won't find me trying to make a positive proof of anything. I believe that deciding on, for example, an optimal taxation policy takes very many numbers and statistical

models and other things which are well beyond the scope of this FAQ, and may well have different answers at different levels and in different areas.

What I want to do in most cases is not prove that the government works better than the free market, or vice versa, but to disprove theories that say we can be absolutely certain free market always works better than government before we even investigate the issue. After that, we may still find that this is indeed one of the cases where the free market works better than the government, but we will have to prove it instead of viewing it as self-evident from first principles.

0.5: Why write a Non-Libertarian FAQ? Isn't statism a bigger problem than libertarianism?

Yes. But you never run into Stalinists at parties. At least not serious Stalinists over the age of twenty-five, and not the interesting type of parties. If I did, I guess I'd try to convince them not to be so statist, but the issue's never come up.

But the world seems positively full of libertarians nowadays. And I see very few attempts to provide a complete critique of libertarian philosophy. There are a bunch of ad hoc critiques of specific positions: people arguing for socialist health care, people in favor of gun control. But one of the things that draws people to libertarianism is that it is a unified, harmonious system. Unlike the mix-and-match philosophies of the Democratic and Republican parties, libertarianism is coherent and sometimes even derived from first principles. The only way to convincingly talk someone out of libertarianism is to launch a challenge on the entire system.

There are a few existing documents trying to do this (see Mike Huben's Critiques of Libertarianism and Mark Rosenfelder's What's (Still) Wrong With Libertarianism for two of the better ones), but I'm not satisfied with any of them. Some of them are good but incomplete. Others use things like social contract theory, which I find nonsensical and libertarians find repulsive. Or they have an

overly rosy view of how consensual taxation is, which I don't fall for and which libertarians *definitely* don't fall for.

The main reason I'm writing this is that I encounter many libertarians, and I need a single document I can point to explaining why I don't agree with them. The existing anti-libertarian documentation makes too many arguments I don't agree with for me to feel really comfortable with it, so I'm writing this one myself. I don't encounter too many Stalinists, so I don't have this problem with them and I don't see any need to write a rebuttal to their position.

If you really need a pro-libertarian FAQ to use on an overly statist friend, Google suggests The Libertarian FAQ.

0.6: How is this FAQ structured?

I've divided it into three main sections. The first addresses some very abstract principles of economics. They may not be directly relevant to politics, but since most libertarian philosophies start with abstract economic principles, a serious counterargument has to start there also. Fair warning: there are people who can discuss economics without it being INCREDIBLY MIND-NUMBINGLY BORING, but I am not one of them.

The second section deals with more concrete economic and political problems like the tax system, health care, and criminal justice.

The third section deals with moral issues, like whether it's ever permissible to initiate force. Too often I find that if I can convince a libertarian that government regulation can be effective, they respond that it doesn't matter because it's morally repulsive, and then once I've finished convincing them it isn't, they respond that it never works anyway. By having sections dedicated to both practical and moral issues, I hope to make that sort of bait-and-switch harder to achieve, and to allow libertarians to evaluate the moral and practical arguments against their position in whatever order they find appropriate.

Part A: Economic Issues

The Argument:

In a free market, all trade has to be voluntary, so you will never agree to a trade unless it benefits you.

Further, you won't make a trade unless you think it's the best possible trade you can make. If you knew you could make a better one, you'd hold out for that. So trades in a free market are not only better than nothing, they're also the best possible transaction you could make at that time.

Labor is no different from any other commercial transaction in this respect. You won't agree to a job unless it benefits you more than anything else you can do with your time, and your employer won't hire you unless it benefits her more than anything else she can do with her money. So a voluntarily agreed labor contract must benefit both parties, and must do so more than any other alternative.

If every trade in a free market benefits both parties, then any time the government tries to restrict trade in some way, it must hurt both parties. Or, to put it another way, you can help someone by giving them more options, but you can't help them by taking away options. And in a free market, where everyone starts with all options, all the government can do is take options away.

The Counterargument:

This treats the world as a series of producer-consumer dyads instead of as a system in which every transaction affects everyone else. Also, it treats consumers as coherent entities who have specific variables like "utility" and "demand" and know exactly what they are, which doesn't always work.

In the remainder of this section, I'll be going over several ways the free market can fail and several ways a regulated market can overcome those failures. I'll focus on four main things: externalities, coordination problems, irrational choice, and lack of information. I did warn you it would be mind-numbingly boring.

1. Externalities

1.1: What is an externality?

An externality is when I make a trade with you, but it has some accidental effect on other people who weren't involved in the trade.

Suppose for example that I sell my house to an amateur wasp farmer. Only he's not a very good wasp farmer, so his wasps usually get loose and sting people all over the neighborhood every couple of days.

This trade between the wasp farmer and myself has benefitted both of us, but it's harmed people who weren't consulted; namely, my neighbors, who are now locked indoors clutching cans of industrial-strength insect repellent. Although the trade was voluntary for both the wasp farmer and myself, it wasn't voluntary for my neighbors.

Another example of externalities would be a widget factory that spews carcinogenic chemicals into the air. When I trade with the widget factory I'm benefitting - I get widgets - and they're benefitting - they get money. But the people who breathe in the carcinogenic chemicals weren't consulted in the trade.

1.2: But aren't there are libertarian ways to solve externalities that don't involve the use of force?

To some degree, yes. You can, for example, refuse to move into any neighborhood unless everyone in town has signed a contract agreeing not to raise wasps on their property.

But getting every single person in a town of thousands of people to sign a contract every time you think of something else you want banned might be a little difficult. More likely, you would want everyone in town to unanimously agree to a contract saying that certain things, which could be decided by some procedure requiring less than unanimity, could be banned from the neighborhood - sort of like the existing concept of neighborhood associations.

But convincing every single person in a town of thousands to join the neighborhood association would be near impossible, and all it would take would be a single holdout who starts raising wasps and all your work is useless. Better, perhaps, to start a new town on your

own land with a pre-existing agreement that before you're allowed to move in you must belong to the association and follow its rules. You could even collect dues from the members of this agreement to help pay for the people you'd need to enforce it.

But in this case, you're not coming up with a clever libertarian way around government, you're just reinventing the concept of government. There's no difference between a town where to live there you have to agree to follow certain terms decided by association members following some procedure, pay dues, and suffer the consequences if you break the rules - and a regular town with a regular civic government.

As far as I know there is no loophole-free way to protect a community against externalities besides government and things that are functionally identical to it.

1.3: Couldn't consumers boycott any company that causes externalities?

Only a small proportion of the people buying from a company will live near the company's factory, so this assumes a colossal amount of both knowledge and altruism on the part of most consumers. See also the general discussion of why boycotts almost never solve problems in the next session.

1.4: What is the significance of externalities?

They justify some environmental, zoning, and property use regulations.

2. Coordination Problems

2.1: What are coordination problems?

Coordination problems are cases in which everyone agrees that a certain action would be best, but the free market cannot coordinate them into taking that action.

As a thought experiment, let's consider aquaculture (fish farming) in a lake. Imagine a lake with a thousand identical fish farms owned by

a thousand competing companies. Each fish farm earns a profit of $1000/month. For a while, all is well.

But each fish farm produces waste, which fouls the water in the lake. Let's say each fish farm produces enough pollution to lower productivity in the lake by $1/month.

A thousand fish farms produce enough waste to lower productivity by $1000/month, meaning none of the fish farms are making any money. Capitalism to the rescue: someone invents a complex filtering system that removes waste products. It costs $300/month to operate. All fish farms voluntarily install it, the pollution ends, and the fish farms are now making a profit of $700/month - still a respectable sum.

But one farmer (let's call him Steve) gets tired of spending the money to operate his filter. Now one fish farm worth of waste is polluting the lake, lowering productivity by $1. Steve earns $999 profit, and everyone else earns $699 profit.

Everyone else sees Steve is much more profitable than they are, because he's not spending the maintenance costs on his filter. They disconnect their filters too.

Once four hundred people disconnect their filters, Steve is earning $600/month - less than he would be if he and everyone else had kept their filters on! And the poor virtuous filter users are only making $300. Steve goes around to everyone, saying "Wait! We all need to make a voluntary pact to use filters! Otherwise, everyone's productivity goes down."

Everyone agrees with him, and they all sign the Filter Pact, except one person who is sort of a jerk. Let's call him Mike. Now everyone is back using filters again, except Mike. Mike earns $999/month, and everyone else earns $699/month. Slowly, people start thinking they too should be getting big bucks like Mike, and disconnect their filter for $300 extra profit…

A self-interested person never has any incentive to use a filter. A self-interested person has some incentive to sign a pact to make

everyone use a filter, but in many cases has a stronger incentive to wait for everyone *else* to sign such a pact but opt out himself. This can lead to an undesirable equilibrium in which no one will sign such a pact.

The most profitable solution to this problem is for Steve to declare himself King of the Lake and threaten to initiate force against anyone who doesn't use a filter. This regulatory solution leads to greater total productivity for the thousand fish farms than a free market could.

The classic libertarian solution to this problem is to try to find a way to privatize the shared resource (in this case, the lake). I intentionally chose aquaculture for this example because privatization doesn't work. Even after the entire lake has been divided into parcels and sold to private landowners (waterowners?) the problem remains, since waste will spread from one parcel to another regardless of property boundaries.

2.1.1: Even without anyone declaring himself King of the Lake, the fish farmers would voluntarily agree to abide by the pact that benefits everyone.

Empirically, no. This situation happens with wild fisheries all the time. There's some population of cod or salmon or something which will be self-sustaining as long as it's not overfished. Fishermen come in and catch as many fish as they can, overfishing it. Environmentalists warn that the fishery is going to collapse. Fishermen find this worrying, but none of them want to fish less because then their competitors will just take up the slack. Then the fishery collapses and everyone goes out of business. The most famous example is the Collapse of the Northern Cod Fishery, but there are many others in various oceans, lakes, and rivers.

If not for resistance to government regulation, the Canadian governments could have set strict fishing quotas, and companies could still be profitably fishing the area today. Other fisheries that do have government-imposed quotas are much more successful.

2.1.2: I bet [extremely complex privatization scheme that takes into account the ability of cod to move across property boundaries and the migration patterns of cod and so on] could have saved the Atlantic cod too.

Maybe, but *left to their own devices, cod fishermen never implemented or recommended that scheme*. If we ban all government regulation in the environment, that won't make fishermen suddenly start implementing complex privatization schemes that they've never implemented before. It will just make fishermen keep doing what they're doing while tying the hands of the one organization that has a track record of actually solving this sort of problem in the real world.

2.2: How do coordination problems justify environmental regulations?

Consider the process of trying to stop global warming. If everyone believes in global warming and wants to stop it, it's still not in any one person's self-interest to be more environmentally conscious. After all, that would make a major impact on her quality of life, but a negligible difference to overall worldwide temperatures. If everyone acts only in their self-interest, then no one will act against global warming, even though stopping global warming is in everyone's self-interest. However, everyone would support the institution of a government that uses force to make *everyone* more environmentally conscious.

Notice how well this explains reality. The government of every major country has publicly declared that they think solving global warming is a high priority, but every time they meet in Kyoto or Copenhagen or Bangkok for one of their big conferences, the developed countries would rather the developing countries shoulder the burden, the developing countries would rather the developed countries do the hard work, and so nothing ever gets done.

The same applies *mutans mutandis* to other environmental issues like the ozone layer, recycling, and anything else where one person

cannot make a major difference but many people acting together can.

2.3: How do coordination problems justify regulation of ethical business practices?

The normal libertarian belief is that it is unnecessary for government to regulate ethical business practices. After all, if people object to something a business is doing, they will boycott that business, either incentivizing the business to change its ways, or driving them into well-deserved bankruptcy. And if people don't object, then there's no problem and the government shouldn't intervene.

A close consideration of coordination problems demolishes this argument. Let's say Wanda's Widgets has one million customers. Each customer pays it $100 per year, for a total income of $100 million. Each customer prefers Wanda to her competitor Wayland, who charges $150 for widgets of equal quality. Now let's say Wanda's Widgets does some unspeakably horrible act which makes it $10 million per year, but offends every one of its million customers.

There is no incentive for a single customer to boycott Wanda's Widgets. After all, that customer's boycott will cost the customer $50 (she will have to switch to Wayland) and make an insignificant difference to Wanda (who is still earning $99,999,900 of her original hundred million). The customer takes significant inconvenience, and Wanda neither cares nor stops doing her unspeakably horrible act (after all, it's giving her $10 million per year, and only losing her $100).

The only reason it would be in a customer's interests to boycott is if she believed over a hundred thousand other customers would join her. In that case, the boycott would be costing Wanda more than the $10 million she gains from her unspeakably horrible act, and it's now in her self-interest to stop committing the act. However, unless each boycotter believes 99,999 others will join her, she is inconveniencing herself for no benefit.

Furthermore, if a customer offended by Wanda's actions believes 100,000 others will boycott Wanda, then it's in the customer's self-interest to "defect" from the boycott and buy Wanda's products. After all, the customer will lose money if she buys Wayland's more expensive widgets, and this is unnecessary – the 100,000 other boycotters will change Wanda's mind with or without her participation.

This suggests a "market failure" of boycotts, which seems confirmed by experience. We know that, despite many companies doing very controversial things, there have been very few successful boycotts. Indeed, few boycotts, successful or otherwise, ever make the news, and the number of successful boycotts seems much less than the amount of outrage expressed at companies' actions.

The existence of government regulation solves this problem nicely. If >51% of people disagree with Wanda's unspeakably horrible act, they don't need to waste time and money guessing how many of them will join in a boycott, and they don't need to worry about being unable to conscript enough defectors to reach critical mass. They simply vote to pass a law banning the action.

2.3.1: I'm not convinced that it's really that hard to get a boycott going. If people really object to something, they'll start a boycott regardless of all that coordination problem stuff.

So, you're boycotting Coke because they're hiring local death squads to kidnap, torture, and murder union members and organizers in their sweatshops in Colombia, right?

Not a lot of people to whom I have asked this question have ever answered "yes". Most of them had never heard of the abuses before. A few of them vaguely remembered having heard something about it, but dismissed it as "you know, multinational corporations do a lot of sketchy things." I've only met one person who's ever gone so far as to walk twenty feet further to get to the Pepsi vending machine.

If you went up to a random guy on the street and said "Hey, does hiring death squads to torture and kill Colombians who protest about

terrible working conditions bother you?" 99.9% of people would say yes. So why the disconnect between words and actions? People could just be lying - they could say they cared so they sounded compassionate, but in reality it doesn't really bother them.

But maybe it's something more complicated. Perhaps they don't have the brainpower to keep track of every single corporation that's doing bad things and just how bad they are. Perhaps they've compartmentalized their lives and after they leave their Amnesty meetings it just doesn't register that they should change their behaviour in the supermarket. Or perhaps the Coke = evil connection is too tenuous and against the brain's ingrained laws of thought to stay relevant without expending extraordinary amounts of willpower. Or perhaps there's some part of the subconscious that really is worry about that game theory and figuring it has no personal incentive to join the boycott.

And God forbid that it's something more complicated than that. Imagine if the company that made the mining equipment that was bought by the mining company that mined the aluminum that was bought by Coke to make their cans was doing something unethical. You think you could convince enough people to boycott Coke that Coke would boycott the mining company that the mining company would boycott the equipment company that the equipment company would stop behaving unethically?

If we can't trust people to stay off Coke when it uses death squads and when Pepsi tastes *exactly the same* (don't argue with me on that one!) how can we assume people's purchasing decisions will always act as a general moral regulatory method for the market?

2.3.2: And you really think governments can do better?

Sure seems that way. Many laws currently exist banning businesses from engaging in unethical practices. Some of these laws were passed by direct ballot. Others were passed by representatives who have incentives to usually follow the will of their consitutents. So it seems fair to say that there are a lot of business practices that more than 51% of people thought should be banned.

But the very fact that a law was needed to ban them proves that those 51% of people weren't able to organize a successful boycott. More than half of the population, sometimes much more, hated some practice so much they thought it should be *illegal*, yet that wasn't enough to provide an incentive for the company to stop doing it until the law took effect.

To me, that confirms that boycotts are a very poor way of allowing people's morals to influence corporate conduct.

2.4: How do coordination problems justify government spending on charitable causes?

Because failure to donate to a charitable cause might also be because of a coordination problem.

How many people want to end world hunger? I've never yet met someone who would answer with a "not me!", but maybe some of those people are just trying to look good in front of other people, so let's make a conservative estimate of 50%.

There's a lot of dispute over what it would mean to "end world hunger", all the way from "buy and ship food every day to everyone who is hungry that day" all the way to "create sustainable infrastructure and economic development such that everyone naturally produces enough food or money". There are various estimates about how much these different definitions would cost, all the way from "about $15 billion a year" to "about $200 billion a year" - permanently in the case of shipping food, and for a decade or two in the case of promoting development.

Even if we take the highest possible estimate, it's still *well* below what you would make if 50% of the population of the world donated $1/week to the cause. Now, certainly there are some very poor people in the world who couldn't donate $1/week, but there are also some very rich people who could no doubt donate much, much more.

So we have two possibilities. Either the majority of people don't care enough about world hunger to give a dollar a week to end it, or

something else is going on.

That something else is a coordination problem. No one expects anyone else to donate a dollar a week, so they don't either. And although somebody could shout very loudly "Hey, let's all donate $1 a week to fight world hunger!" no one would expect anyone else to listen to that person, so they wouldn't either.

When the government levies tax money on everyone in the country and then donates it to a charitable cause, it is often because everyone in the country supports that charitable cause but a private attempt to show that support would fall victim to coordination problems.

2.5: How do coordination problems justify labor unions and other labor regulation?

It is frequently proposed that workers and bosses are equal negotiating partners bargaining on equal terms, and only the excessive government intervention on the side of labor that makes the negotiating table unfair. After all, both need something from one another: the worker needs money, the boss labor. Both can end the deal if they don't like the terms: the boss can fire the worker, or the worker can quit the boss. Both have other choices: the boss can choose a different employee, the worker can work for a different company. And yet, strange to behold, having proven the fundamental equality of workers and bosses, we find that everyone keeps acting as if bosses have the better end of the deal.

During interviews, the prospective employee is often nervous; the boss rarely is. The boss can ask all sorts of things like that the prospective pay for her own background check, or pee in a cup so the boss can test the urine for drugs; the prospective employee would think twice before daring make even so reasonable a request as a cup of coffee. Once the employee is hired, the boss may ask on a moment's notice that she work a half hour longer or else she's fired, and she may not dare to even complain. On the other hand, if she were to so much as ask to be allowed to start work thirty minutes later to get more sleep or else she'll quit, she might well be laughed out of the company. A boss may, and very often does, yell at an

employee who has made a minor mistake, telling her how stupid and worthless she is, but rarely could an employee get away with even politely mentioning the mistake of a boss, even if it is many times as unforgivable.

The naive economist who truly believes in the equal bargaining position of labor and capital would find all of these things very puzzling.

Let's focus on the last issue; a boss berating an employee, versus an employee berating a boss. Maybe the boss has one hundred employees. Each of these employees only has one job. If the boss decides she dislikes an employee, she can drive her to quit and still be 99% as productive while she looks for a replacement; once the replacement is found, the company will go on exactly as smoothly as before.

But if the employee's actions drive the boss to fire her, then she must be completely unemployed until such time as she finds a new job, suffering a long period of 0% productivity. Her new job may require a completely different life routine, including working different hours, learning different skills, or moving to an entirely new city. And because people often get promoted based on seniority, she probably won't be as well paid or have as many opportunities as she did at her old company. And of course, there's always the chance she won't find another job at all, or will only find one in a much less tolerable field like fast food.

We previously proposed a symmetry between a boss firing a worker and a worker quitting a boss, but actually they could not be more different. For a boss to fire a worker is at most a minor inconvenience; for a worker to lose a job is a disaster. The Holmes-Rahe Stress Scale, a measure of the comparative stress level of different life events, puts being fired at 47 units, worse than the death of a close friend and nearly as bad as a jail term. Tellingly, "firing one of your employees" failed to make the scale.

This fundamental asymmetry gives capital the power to create more asymmetries in its favor. For example, bosses retain a level of

control on workers even after they quit, because a worker may very well need a letter of reference from a previous boss to get a good job at a new company. On the other hand, a prospective employee who asked her prospective boss to produce letters of recommendation from her previous workers would be politely shown the door; we find even the image funny.

The proper level negotiating partner to a boss is not one worker, but all workers. If the boss lost *all* workers at once, then she would be at 0% productivity, the same as the worker who loses her job. Likewise, if *all* the workers approached the boss and said "We want to start a half hour later in the morning or we all quit", they might receive the same attention as the boss who said "Work a half hour longer each day or you're all fired".

But getting all the workers together presents coordination problems. One worker has to be the first to speak up. But if one worker speaks up and doesn't get immediate support from all the other workers, the boss can just fire that first worker as a troublemaker. Being the first worker to speak up has major costs - a good chance of being fired - but no benefits - all workers will benefit equally from revised policies no matter who the first worker to ask for them is.

Or, to look at it from the other angle, if only one worker sticks up for the boss, then intolerable conditions may well still get changed, but the boss will remember that one worker and maybe be more likely to promote her. So even someone who hates the boss's policies has a strong selfish incentive to stick up for her.

The ability of workers to coordinate action without being threatened or fired for attempting to do so is the only thing that gives them any negotiating power at all, and is necessary for a healthy labor market. Although we can debate the specifics of exactly how much protection should be afforded each kind of coordination, the fundamental principle is sound.

2.5.1: But workers don't need to coordinate. If working conditions are bad, people can just change jobs, and that would solve the bad conditions.

About three hundred Americans commit suicide for work-related reasons every year - this number doesn't count those who attempt suicide but fail. The reasons cited by suicide notes, survivors and researchers investigating the phenomenon include on-the-job bullying, poor working conditions, unbearable hours, and fear of being fired.

I don't claim to understand the thought processes that would drive someone to do this, but given the rarity and extremity of suicide, we can assume for every worker who goes ahead with suicide for work-related reasons, there are a hundred or a thousand who feel miserable but not quite suicidal.

If people are literally killing themselves because of bad working conditions, it's safe to say that life is more complicated than the ideal world in which everyone who didn't like their working conditions quits and get a better job elsewhere (see the next section, Irrationality).

I note in the same vein stories from the days before labor regulations when employers would ban workers from using the restroom on jobs with nine hour shifts, often ending in the workers wetting themselves. This seems like the sort of thing that provides so much humiliation to the workers, and so little benefit to the bosses, that a free market would eliminate it in a split second. But we know that it was a common policy in the 1910s and 1920s, and that factories with such policies never wanted for employees. The same is true of factories that literally locked their workers inside to prevent them from secretly using the restroom or going out for a smoking break, leading to disasters like the Triangle Shirtwaist Fire when hundreds of workers died when the building they were locked inside burnt down. And yet even after this fire, the practice of locking workers inside buildings only stopped when the government finally passed regulation against it.

3. Irrational Choices

3.1: What do you mean by "irrational choices"?

A company (Thaler, 2007, download study as .pdf) gives its employees the opportunity to sign up for a pension plan. They contribute a small amount of money each month, and the company will also contribute some money, and overall it ends up as a really good deal for the employees and gives them an excellent retirement fund. Only a small minority of the employees sign up.

The libertarian would answer that this is fine. Although some outsider might condescendingly declare it "a really good deal", the employees are the most likely to understand their own unique financial situation. They may have a better pension plan somewhere else, or mistrust the company's promises, or expect not to need much money in their own age. For some outsider to declare that they are *wrong* to avoid the pension plan, or worse to try to *force* them into it for their own good, would be the worst sort of arrogant paternalism, and an attack on the employees' dignity as rational beings.

Then the company switches tactics. It automatically signs the employees up for the pension plan, but offers them the option to opt out. This time, only a small minority of the employees opt out.

That makes it very hard to spin the first condition as the employees rationally preferring not to participate in the pension plan, since the second condition reveals the opposite preference. It looks more like they just didn't have the mental energy to think about it or go through the trouble of signing up. And in the latter condition, they didn't have the mental energy to think about it or go through the trouble of opting out.

If the employees were rationally deciding whether or not to sign up, then some outsider regulating their decision would be a disaster. But if the employees are making demonstrably irrational choices because of a lack of mental energy, and if people do so consistently and predictably, then having someone else who has considered the issue in more depth regulate their choices could lead to a better outcome.

3.1.1: So what's going on here?

Old-school economics assumed choice to be “revealed preference”: an individual’s choices will invariably correspond to their preferences, and imposing any other set of choices on them will result in fewer preferences being satisfied.

In some cases, economists have gone to absurd lengths to defend this model. For example, Bryan Caplan says that when drug addicts say they wish that they could quit drugs, they must be lying, since they haven’t done so. Seemingly unsuccessful attempts to quit must be elaborate theater, done to convince other people to continue supporting them, while they secretly enjoy their drugs as much as ever.

But the past fifty years of cognitive science have thoroughly demolished this “revealed preference” assumption, showing that people’s choices result from a complex mix of external compulsions, internal motivations, natural biases, and impulsive behaviors. These decisions usually approximate fulfilling preferences, but sometimes they fail in predictable and consistent ways. The field built upon these insights is called “behavioral economics”, and you can find more information in books like Judgment Under Uncertainty, Cognitive Illusions, and Predictably Irrational, or on the website Less Wrong.

3.2: Why does this matter?

The gist of this research, as it relates to the current topic, is that people don’t always make the best choice according to their preferences. Sometimes they consistently make the easiest or the most superficially attractive choice instead. It may be best not to think of them as a “choice” at all, but as a reflexive reaction to certain circumstances, which often but not always conforms to rationality.

Such possibilities cast doubt on the principle that every trade that can be voluntarily made should be voluntarily made.

If people’s decisions are not randomly irrational, but systematically irrational in predictable ways, that raises the possibility that people

who are aware of these irrationalities may be able to do better than the average person in particular fields where the irrationalities are more common, raising the possibility that paternalism can sometimes be justified.

3.2.1: Why should the government protect people from their own irrational choices?

By definition of "irrational", people will be happier and have more of their preferences satisfied if they do not make irrational choices. By the principles of the free market, as people make more rational decisions the economy will also improve.

If you mean this question in a *moral* sense, more like "How *dare* the government presume to protect me from my own irrational choices!", see the section on Moral Issues.

3.2.2: What is the significance of predictably irrational behavior?

It justifies government-mandated pensions, some consumer safety and labor regulations, advertising regulations, concern about addictive drugs, and public health promotion, among other things.

4. Lack of Information

4.1: What do you mean by "lack of information"?

Many economic theories start with the assumption that everyone has perfect information about everything. For example, if a company's products are unsafe, these economic theories assume consumers know the product is unsafe, and so will buy less of it.

No economist literally believes consumers have perfect information, but there are still strong arguments for keeping the "perfect information" assumption. These revolve around the idea that consumers will be motivated to pursue information about things that are important to them. For example, if they care about product safety, they will fund investigations into product safety, or only buy products that have been certified safe by some credible third party. The only case in which a consumer would buy something without

information on it is if the consumer had no interest in the information, or wasn't willing to pay as much for the information as it would cost, in which case the consumer doesn't care much about the information anyway, and it is a success rather than a failure of the market that it has not given it to her.

In nonlibertarian thought, people care so much about things like product safety and efficacy, or the ethics of how a product is produced, that the government needs to ensure them. In libertarian thought, if people really care about product safety, efficacy and ethics, the market will ensure them itself, and if they genuinely don't care, that's okay too.

4.1.1: And what's wrong with the libertarian position here?

Section 5 describes how we can sometimes predict when people will make irrational choices. One of the most consistent irrational choices people make is buying products without spending as much effort to gather information as the amount they care about these things would suggest. So in fact, the nonlibertarians are right: if there were no government regulation, people who care a lot about things like safety and efficacy would consistently be stuck with unsafe and ineffective products, and the market would not correct these failures.

4.2: Is this really true? Surely people would investigate the safety, ethics, and efficacy of the products they buy.

Below follows a list of statements about products. Some are real, others are made up. Can you identify which are which?

1. Some processed food items, including most Kraft cheese products, contain methylarachinate, an additive which causes a dangerous anaphylactic reaction in 1/31000 people who consume it. They have been banned in Canada, but continue to be used in the United States after intense lobbying from food industry interests.

2. Commonly used US-manufactured wood products, including almost all plywood, contain formaldehyde, a compound known to cause cancer. This has been known in scientific circles for years, but

was only officially reported a few months ago because of intense chemical industry lobbying to keep it secret. Formaldehyde-containing wood products are illegal in the EU and most other developed nations.

3. Total S.A., an oil company that owns fill-up stations around the world, sometimes uses slave labor in repressive third-world countries to build its pipelines and oil wells. Laborers are coerced to work for the company by juntas funded by the corporation, and are shot or tortured if they refuse. The company also helps pay for the military muscle needed to keep the juntas in power.

4. Microsoft has cooperated with the Chinese government by turning over records from the Chinese equivalents of its search engine "Bing" and its hotmail email service, despite knowing these records would be used to arrest dissidents. At least three dissidents were arrested based on the information and are currently believed to be in jail or "re-education" centers.

5. Wellpoint, the second largest US health care company, has a long record of refusing to provide expensive health care treatments promised in some of its plans by arguing that their customers have violated the "small print" of the terms of agreement; in fact they make it so technical that almost all customers violate them unknowingly, then only cite the ones who need expensive treatment. Although it has been sued for these practices at least twice, both times it has used its legal muscle to tie the cases up in court long enough that the patients settled for an undisclosed amount believed to be fraction of the original benefits promised.

6. Ultrasonic mosquito repellents like those made by GSI, which claim to mimic frequencies produced by the mosquito's natural predator, the bat, do not actually repel mosquitoes. Studies have shown that exactly as many mosquitoes inhabit the vicinity of such a mosquito repellent as anywhere else.

7. Listerine (and related mouth washes) probably do not eliminate bad breath. Although it may be effective at first, in the long term it generally increases bad breath by drying out the mouth and

inhibiting the salivary glands. This may also increase the population of dental bacteria. Most top dentists recommend avoiding mouth wash or using it very sparingly.

8. The most popular laundry detergents, including most varieties of Tide and Method, have minimal to zero ability to remove stains from clothing. They mostly just makes clothing smell better when removed from the laundry. Some of the more expensive alkylbenzenesulfonate detergents have genuine stain-removing action, but aside from the cost, these detergents have very strong smells and are unpopular.

4.2.1: Okay, I admit I'm not sure of most of these. What's your point?

This is a complicated FAQ about complicated philosophical issues. Most likely its readers are in the top few percentiles in terms of intelligence and education.

And we live in a world where there are many organizations, both private and governmental, that exist to evaluate products and disseminate information about their safety.

And all of the companies and products above are popular ones that most American consumers have encountered and had to make purchasing decisions about. I tried to choose safety issues that were extremely serious and carried significant risks of death, and ethical issues involving slavery and communism, which would be of particular importance to libertarians.

If the test was challenging, it means that the smartest and best-educated people in a world full of consumer safety and education organizations don't bother to look up important life-or-death facts specifically tailored to be relevant to them about the most popular products and companies they use every day.

And if that's the case, why would you believe that less well-educated people in a world with less consumer safety information trying to draw finer distinctions between more obscure products will

definitely seek out the consumer information necessary allows them to avoid unsafe, unethical, or ineffective products?

The above test is an attempt at experimental proof that people don't seek out even the product information that is genuinely important to them, but instead take the easy choice of buying whatever's convenient based on information they get from advertising campaigns and the like.

4.2.2: Fine, fine, what are the answers to the test?

Four of them are true and four of them are false, but I'm not saying which are which, in the hopes that people will observe their own thought processes when deciding whether or not it's worth looking up.

4.2.3: Right, well of course people don't look up product information *now* because the government regulates that for them. In a real libertarian society, they would be more proactive.

All of the four true items on the test above are true *in spite* of government regulation. Clearly, there are still significant issues even in a regulated environment.

If you honestly believe you have no incentive to look up product information because you trust the government to take care of that, then you're about ten times more statist than I am, and *I'm the guy writing the Non-Libertarian FAQ*.

4.3: What other unexpected consequences might occur without consumer regulation?

It could destroy small business.

In the absence of government regulation, you would have to trust corporate self-interest to regulate quality. And to some degree you can do that. Wal-Mart and Target are both big enough and important enough that if they sold tainted products, it would make it into the newspaper, there would be a big outcry, and they would be forced to stop. One could feel quite safe shopping at Wal-Mart.

But suppose on the way to Wal-Mart, you see a random mom-and-pop store that looks interesting. What do you know about its safety standards? Nothing. If they sold tainted or defective products, it would be unlikely to make the news; if it were a small enough store, it might not even make the Internet. Although you expect the CEO of Wal-Mart to be a reasonable man who understands his own self-interest and who would enforce strict safety standards, you have no idea whether the owner of the mom-and-pop store is stupid, lazy, or just assumes (with some justification) that no one will ever notice his misdeeds. So you avoid the unknown quantity and head to Wal-Mart, which you know is safe.

Repeated across a million people in a thousand cities, big businesses get bigger and small businesses get unsustainable.

4.4: What is the significance of lack of information?

It justifies some consumer and safety regulations, and the taxes necessary to pay for them.

Part B: Social Issues

The Argument:

Those who work hardest (and smartest) should get the most money. Not only should we not begrudge them that money, but we should thank them for the good they must have done for the world in order to satisfy so many consumers.

People who do not work hard should not get as much money. If they want more money, they should work harder. Getting more money without working harder or smarter is unfair, and indicative of a false sense of entitlement.

Unfortunately, modern liberal society has internalized the opposite principle: that those who work hardest are greedy people who must have stolen from those who work less hard, and that we should distrust them at until they give most of their ill-gotten gains away to others. The "progressive" taxation system as it currently exists serves this purpose. This way of thinking is not only morally wrong-

headed, but economically catastrophic. Leaving wealth in the hands of the rich would "make the pie bigger", allowing the extra wealth to "trickle down" to the poor naturally.

The Counterargument:

Hard work and intelligence are contributory factors to success, but depending on the way you phrase the question, you find you need other factors to explain between one-half and nine-tenths of the difference in success within the United States; within the world at large the numbers are much higher.

If we think factors other than hard work and intelligence determining success are "unfair", then most of Americans' life experiences are determined by "unfair" factors.

Although it would be overly ambitious to want to completely eliminate all unfairness, we know that most other developed countries have successfully eliminated many of the most glaring types of unfairness, and reaped benefits greater than the costs from doing so.

The progressive tax system is part of this policy of eliminating unfairness, but if you disagree with that, that's okay, as more and more of the country's wealth is staying in the hands of the super-rich. None of this wealth has trickled down to the poor and none of it ever will, as the past thirty years of economic history have repeatedly and decisively demolished the "trickle-down" concept.

None of this implies that any particular rich person is "greedy", whatever that would mean.

5. Just Desserts and Social Mobility

5.1: Government is the recourse of "moochers", who want to take the money of productive people and give it to the poor. But rich people earned their money, and poor people had the chance to earn money but did not. Therefore, the poor do not deserve rich people's money.

The claim of many libertarians is that the wealthy earned their money by the sweat of their brow, and the poor are poor because they did not. The counterclaim of many liberals is that the wealthy gained their wealth by various unfair advantages, and that the poor never had a chance. These two conflicting worldviews have been the crux of many an Internet flamewar.

Luckily, this is an empirical question, and can be solved simply by collecting the relevant data. For example, we could examine whether the children of rich parents are more likely to be rich than poor parents, and, if so, how much more likely they are. This would give us a pretty good estimate of how much of rich people's wealth comes from superior personal qualities, as opposed to starting with more advantages.

If we define "rich" as "income in the top 5%" and "poor" as "income in the bottom 5%" then children of rich parents are about twenty times more likely to become rich themselves than children of poor parents.

But maybe that's an extreme case. Instead let's talk about "upper class" (top 20%) and "lower class" (bottom 20%). A person born to a lower-class family only has a fifty-fifty chance of ever breaking out of the lower class (as opposed to 80% expected by chance), and only about a 3% chance of ending up in the upper class (as opposed to 20% expected by chance). The children of upper class parents are six times more likely to end up in the upper class than the lower class; the children of lower class families are four times more likely to end up in the lower class than the upper class.

The most precise way to measure this question is via a statistic called "intergenerational income mobility", which studies have estimated at between .4 and .6. This means that around half the difference in people's wealth, maybe more, can be explained solely by who their parents are.

Once you add in all the other factors besides how hard you work - like where you live (the average Delawarean earns $30000; the average Mississippian $15000) and the quality of your local school

district, there doesn't seem to be much room for hard work to determine more than about a third of the difference between income.

5.1.1: The conventional wisdom among libertarians is completely different. I've heard of a study saying that people in the lower class are more likely to end up in the upper class than stay in the lower class, even over a period as short as ten years!

First of all, note that this is insane. Since the total must add up to 100%, this would mean that starting off poor actually makes you more likely to end up rich than someone who didn't start off poor. If this were true, we should all send our children to school in the ghetto to maximize their life chances. This should be a red flag.

And, in fact, it is false. Most of the claims of this sort come from a single discredited study. The study focused on a cohort with a median age of twenty-two, then watched them for ten years, then compared the (thirty-two year old) origins with twenty-two year olds, then claimed that the fact that young professionals make more than college students was a fact about social mobility. It was kind of weird.

Why would someone do this? Far be it from me to point fingers, but Glenn Hubbard, the guy who conducted the study, worked for a conservative think tank called the "American Enterprise Institute". You can see a more complete criticism of the study here.

5.1.2: Okay, I acknowledge that at least half of the differences in wealth can be explained by parents. But that needn't be rich parents leaving trust funds to their children. It could also be parents simply teaching their children better life habits. It could even be genes for intelligence and hard work.

This may explain a small part of the issue, but see 5.1.3 and 5.1.3.1, which show that under different socioeconomic conditions, this number markedly decreases. These socioeconomic changes would not be expected to affect things like genetics.

5.1.3: So maybe children of the rich do have better opportunities, but that's life. Some people just start with

advantages not available to others. There's no point in trying to use Big Government to regulate away something that's part of the human condition.

This lack of social mobility isn't part of the human condition, it's a uniquely American problem. Of eleven developed countries investigated in a recent study on income mobility, America came out tenth out of eleven. Their calculation of US intergenerational income elasticity (the number previously cited as probably between .4 and .6) was .47. But other countries in the study had income elasticity as low as .15 (Denmark), .16 (Australia), .17 (Norway), and .19 (Canada). In each of those countries, the overwhelming majority of wealth is earned by hard work rather than inherited.

The United States, is just particularly bad at this; the American Dream turns out to be the "nearly every developed country except America" Dream.

5.1.3.1: That's depressing, but don't try to turn it into a political narrative. Given the government's incompetence and wastefulness, there's no reason to think more government regulation and spending could possibly improve social mobility at all.

Studies show that increasing government spending significantly improves social mobility. States with higher government spending have about 33% more social mobility than states with lower spending.

This also helps explain why other First World countries have better social mobility than we do. Poor American children have very few chances to go to Harvard or Yale; poor Canadian children have a much better chance to go to to UToronto or McGill, where most of their tuition is government-subsidized.

5.2: Then perhaps it is true that rich children start out with a major unfair advantage. But this advantage can be overcome. Poor children may have to work harder than rich children to become rich adults, but this is still *possible*, and so it is still true,

in the important sense, that if you are not rich it's mostly your own fault.

Several years ago, I had an interesting discussion with an evangelical Christian on the ethics of justification by faith. I promise you this will be relevant eventually.

I argued that it is unfair for God to restrict entry to Heaven to Christians alone. After all, 99% of native-born Ecuadorans are Christian, but less than 1% of native born Saudis are same. It follows that the chance of any native-born Ecuadorian of becoming Christian is 99%, and that of any native born Saudi, 1%. So if God judges people by their religion, then within 1% He's basically just decided it's free entry for Ecuadorians, but people born in Saudi Arabia can go to hell (literally).

My Christian friend argued that is not so: that there is a great difference between 0% of Saudis and 1% of Saudis. I answered that no, there was a 1% difference. But he said this 1% proves that the Saudis had free will: that even though all the cards were stacked against them, a few rare Saudis could still choose Christianity.

But what does it mean to have free will, if external circumstances can make 99% of people with free will decide one way in Ecuador, and the opposite way in Saudi Arabia?

I do sort of believe in free will, or at least in "free will". But where my friend's free will was unidirectional, an arrow pointing from MIND to WORLD, my idea of free will is circular: MIND affects WORLD affects MIND affects WORLD and so on.

Yes, it is ultimately the mind and nothing else that decides whether to accept or reject Islam or Christianity. But it is the world that shapes the mind before it does its accepting or rejecting. A man raised in Saudi Arabia uses a mind forged by Saudi culture to make the decision, and chooses Islam. A woman raised in Ecuador uses a mind forged by Ecuador to make the decision, and chooses Christianity. And so there is no contradiction in the saying that the decision between Islam and Christianity is up entirely to the

individual, yet that it is almost entirely culturally determined. For the mind is a box, filled with genes and ideas, and although it is a wonderful magical box that can take things and combine them and forge them into something quite different and unexpected, it is not infinitely magical, and it cannot create out of thin air.

Returning to the question at hand, every poor person has the opportunity to work hard and eventually become rich. Whether that poor person grasps the opportunity comes from that person's own personality. And that person's own personality derives eventually from factors outside that person's control. A clear look at the matter proves it must be so, or else personality would be self-created, like the story of the young man who received a gift of a time machine from a mysterious aged stranger, spent his life exploring past and future, and, in his own age, goes back and gives his time machine to his younger self.

5.2.1: And why is this relevant to politics?

Earlier, I offered a number between .4 and .6 as the proportion of success attributable solely to one's parents' social class. This bears on, but does not wholly answer, a related question: what percentage of my success is my own, and what percentage is attributable to society? People have given answers to this question as diverse as (100%, 0%), (50%, 50%), (0%, 100%).

I boldly propose a different sort of answer: (80%, 100%). Most of my success comes from my own hard work, and all of my own hard work comes from external factors.

If all of our success comes from external factors, then it is reasonable to ask that we "pay it forward" by trying to improve the external factors of others, turning them into better people who will be better able to seize the opportunities to succeed. This is a good deal of the justification for the liberal program of redistribution of wealth and government aid to the poor.

5.2.2: This is all very philosophical. Can you give some concrete examples?

Lead poisoning, for example. It's relatively common among children in poorer areas (about 7% US prevalence) and was even more common before lead paint and leaded gasoline was banned (still >30% in many developing contries).

For every extra ten millionths of a gram per deciliter concentration of lead in their blood, children permanently lose five IQ points; there's a difference of about ten IQ points among children who grew up in areas with no lead at all, and those who grew up in areas with the highest level of lead currently considered "safe". Although no studies have been done on severely lead poisoned children from the era of leaded gasoline, they may have lost twenty or more IQ points from chronic lead exposure.

Further, lead also decreases behavioral inhibition, attention, and self-control. For every ten ug/dl lead increase, children were 50% more likely to have recognized behavioral problems. People exposed to higher levels of blood lead as a child were almost 50% more likely to be arrested for criminal behavior as adults (adjusting for confounders).

Economic success requires self-control, intelligence, and attention. It is cruel to blame people for not seizing opportunities to rise above their background when that background has damaged the very organ responsible for seizing opportunities. And this is why government action, despite a chorus of complaints from libertarians, banned lead from most products, a decision which is (controversially) credited with the most significant global drop in crime rates in decades, but which has certainly contributed to social mobility and opportunity for children who would otherwise be too lead-poisoned to succeed.

Lead is an interesting case because it has obvious neurological effects preventing success. The ability of psychologically and socially toxic environments to prevent success is harder to measure but no less real.

If a poor person can't keep a job solely because she was lead-poisoned from birth until age 16, is it still fair to blame her for her failure? And is it still so unthinkable to take a little bit of money

from everyone who was lucky enough to grow up in an area without lead poisoning, and use it to help her and detoxify her neighborhood?

5.3: What is the significance of whether success is personally or environmentally determined?

It provides justification for redistribution of wealth, and for engineering an environment in which more people are able to succeed.

6. Taxation

6.1: Isn't taxation, the act of taking other people's money by force, inherently evil?

See the Moral Issues section for a more complete discussion of this point.

6.2: Isn't progressive taxation, the tendency to tax the rich at higher rates than the poor, unfair?

The most important justification for progressive tax rates is the idea of marginal utility.

This is easier to explain with movie tickets than money. Suppose different people are alloted a different number of non-transferrable movie tickets for a year; some people get only one, other people get ten thousand.

A person with only two movie ticket might love to have one extra ticket. Perhaps she is a huge fan of X-Men, Batman and Superman, and with only two movie ticket she will only be able to see two of the three movies she's super-excited about this year.

A person with ten movie tickets would get less value from an extra ticket. She can already see the ten movies that year she's most interested in. If she got an eleventh, she'd use it for a movie she might find a bit enjoyable, but it wouldn't be one of her favorites.

A person with a hundred movie tickets would get minimal value from an extra ticket. Even if your tickets are free, you're not likely

to go to the movies a hundred times a year. And even if you did, you'd start scraping the bottom of the barrel in terms of watchable films.

A person with a thousand tickets would get practically no value from an extra ticket. At this point,t here's no way she can go to any more movies. The extra ticket might not have literally zero value - she could burn it for warmth, or write memos on the back of it - but it's pretty worthless.

So although all movie tickets provide an equal service - seeing one movie - one extra movie ticket represents a different amount of value to the person with two tickets and the person with a thousand tickets. Furthermore, 50% of their movie ticket holdings represent a different value to the person with two tickets and the person with a thousand movie tickets. The person with two tickets loses the ability to watch the second-best film of the year. The person with a thousand tickets still has five hundred tickets left, more than enough to see all the year's best films, and at worst will have to buy some real memo paper.

Money works similarly to movie tickets. Your first hundred dollars determine whether you live or starve to death. Your next five hundred dollars determine whether you have a roof over your head or you're freezing out on the street. But by your ten billionth dollar, all you're doing is buying a slightly larger yacht.

50% of what a person with $10,000 makes is more valuable to her than 50% of what a billionaire makes is to the billionaire.

Progressive taxation is an attempt to tax everyone equally, not by lump sum or by percentage, but by *burden*. Just as taking extra movie tickets away from the person with a thousand is more fair than taking some away from the person with only two, so we tax the rich at a higher rate because a proportionate amount of money has less marginal value to them.

6.2.1: But the progressive tax system is unfair and perverse. Imagine the tax rate on people making $100,000 or less is 30%,

and the tax rate on people making more than $100,000 is 50%. You make $100,000, and end up with after tax income of $70,000. Then one day your boss tells you that you did a good job, and gives you a $1 bonus. Now you make $100,001, but end up with only $50,000.50 after tax income. How is that at all fair?

It's not, but this isn't how the tax system works.

What those figures mean is that your first $100,000, no matter how much you earn, is taxed at 30%. Then the money you make *after* that is taxed at 50%. So if you made $100,001, you would be taxed 30% on the first $100,000 (giving you $70,000), and 50% on the next $1 (giving you $.50), for an after-tax income of $70,000.50. The intuitive progression where someone who makes more money ends up with more after-tax income is preserved.

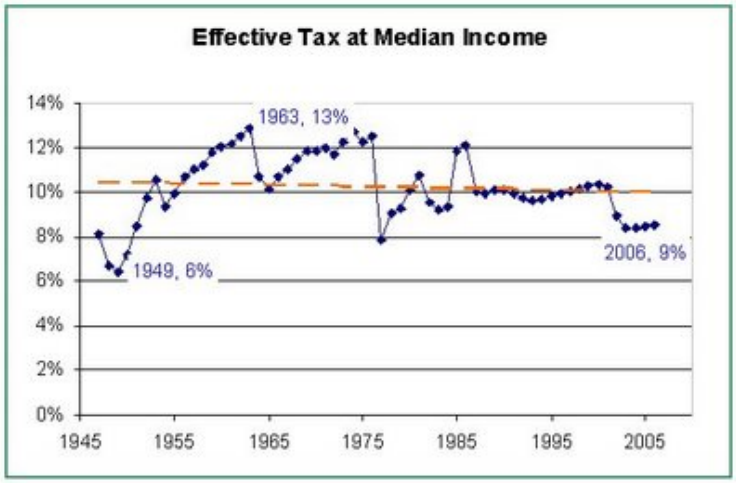
I know most libertarians don't make this mistake, and that there are much stronger arguments against progressive taxation, but this has come up enough times that I thought it was worth mentioning, with apologies to those readers whose time it has wasted.

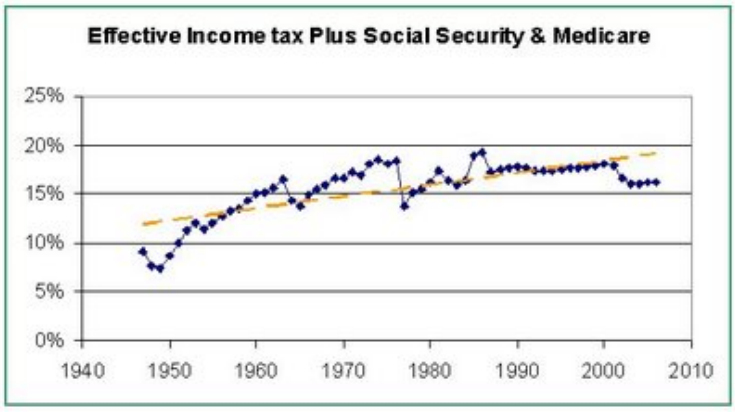
6.3: Taxes are too high.

Too high by what standard?

6.3.1: Too high by historical standards. Thanks to the unstoppable growth of big government, people have to pay more taxes now than ever before.

Actually, income tax rates for people on median income are around the lowest they've been in the past seventy-five years





6.3.1.1: I meant for the rich. It's only tolerable for people on median income because "progressive" governments are squeezing every last dollar out of successful people.

Actually, income tax rates for the rich are around the lowest they've been in the past seventy-five years.

6.3.1.1.1: But I heard that the share of tax revenue coming from the rich is at its highest level ever.

This is true. As the rich get richer and the poor get poorer (see 3.4), more of the money concentrates in the hands of the rich, and so more of the taxes come from the rich as well. This doesn't contradict the point that the tax rates on the rich are near historic lows. **6.3.1.2: I meant for corporations.**

Actually, income tax rates for corporations are around the lowest they've been in the past seventy-five years.

6.3.2: I meant income taxes are too high compared to what's best for the economy, and even best for the Treasury. With taxes as

high as they are, people will stop producing, rather than see so much of each dollar they make go to the government. This will hurt the economy *and* lower tax revenue.

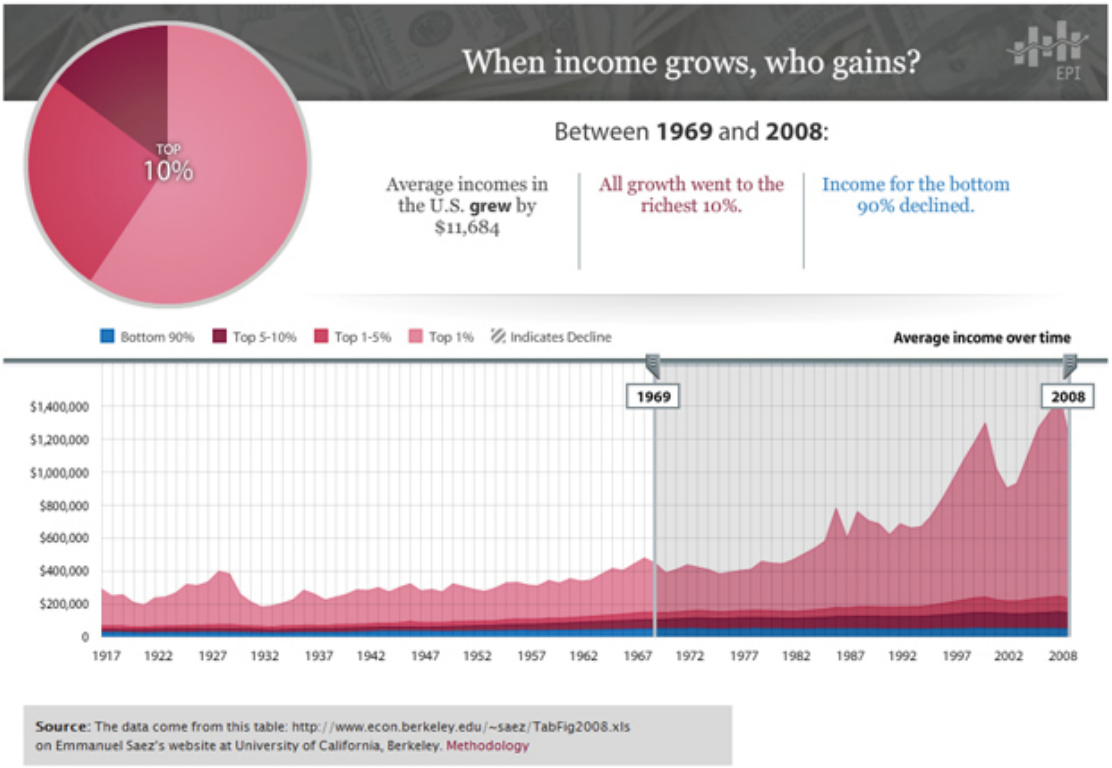
The Laffer curve certainly exists, but the consensus is that we're still well on the left half of it.

Although it's become a truism that high tax rates discourage production, studies have found this to be mostly false, with low elasticity of real income - see for example Gruber & Saez and Saez, Slemrod, and Giertz.

What studies *have* found is a high elasticity of taxable income. That is, raising taxes encourages people to find more tax loopholes, decreasing revenue. However, although this effect means a 10% higher tax rate would lead to less than 10% higher government income, the change in government income would still be positive - even by this stricter criterion, we're still on the left side of the Laffer curve. And of course, this effect could be eliminated by switching to a flat tax or closing tax loopholes.

6.4: Our current tax system is overzealous in its attempts to redistribute money from the rich to the poor. If instead we lowered taxes on the rich, this money would "trickle down" to the rest of the economy, driving growth. Instead of redistributing the pie, we'd make the pie larger for everyone.

If we're in an overzealous campaign for "equality" intended to lower the rich to the level of the poor, we're certainly not doing a very good job of it. Over the past thirty years, the rich have consistently gotten richer. None of this money has trickled down to the poor or middle-class, whose income has remained the same in real terms.



"Trickle-down" should be rejected as an interesting and plausible-sounding economic theory which empirical data have soundly disconfirmed.

6.5: Raising taxes would be useless for the important things like cutting the deficit. The deficit is $1.2 trillion. The most we could realistically raise from extra taxes on the rich would be maybe $200 billion. The most we could raise from insane levels of extra taxes on the rich *and* middle class would be about $500 billion - less than half the deficit. The real problem is spending.

Yes and no.

The deficit is, indeed, *very, very* large. It's so large that no politically palatable option is likely to make more than a small dent in it. This is true of tax increases. It's also true of spending cuts.

Cutting *all* redistributive government services for the poor including welfare, unemployment insurance, disability, food stamps, scholarships, you name it - would save about $200 billion. That's less than 20% of the deficit. Cutting *all* health care, including Medicaid for senior citizens, would only eliminate $400 billion or

so. Even eliminating the entire military down to the last Jeep would only get us $800 billion or so. The targets for cuts that have actually been raised are rounding errors: the Republicans trumpeted an end for government aid to NPR, but this is about $4 million - all of .000003% of the problem.

So “darnit, this one thing doesn’t completely solve the deficit” is not a good reason to reject a proposal. Solving the deficit will, if it’s possible at all, take a lot of different methods, including some unpalatable to liberals, some unpalatable to conservatives, and yes, some unpalatable to libertarians.

In particular, we need to avoid the “bee sting” fallacy, where we have so many problems that we just stop worrying. It would be irresponsible to say that since a few billion dollars doesn’t affect the deficit either way, we might as well just spend $5 billion on some random project we don’t need. For the same reason, it would be irresponsible to say we might as well just renew tax cuts on the rich that cost hundreds of billions of dollars each year.

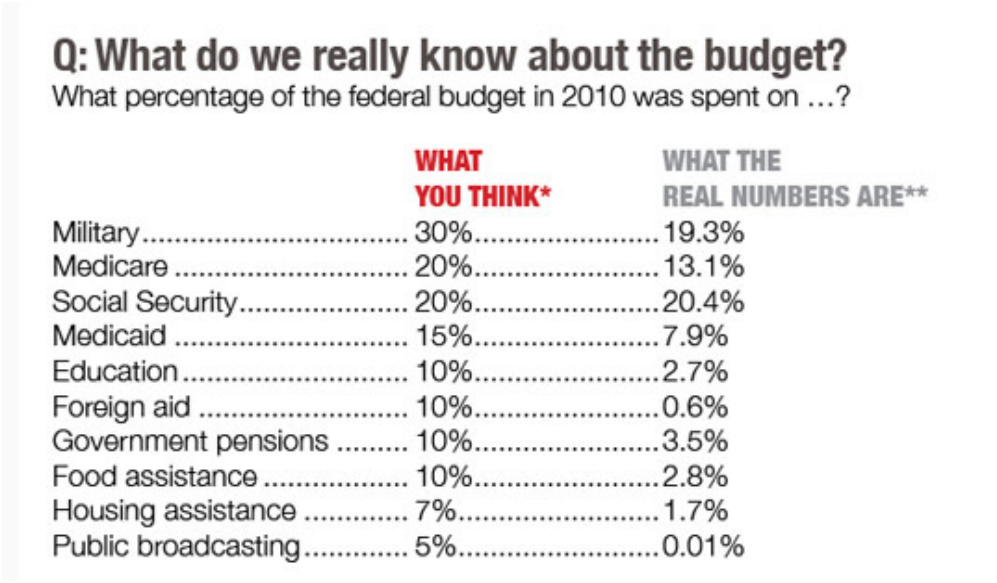
6.6: Taxes are basically a racket where they take my money and then give it to foreign governments and poor people.

According to a CNN poll, on average Americans estimate that about 10% of our taxes go to foreign aid. The real number is about 0.6%.

And although people believe that food and housing for the poor take up about 20% of the federal budget, the real number is actually less than 5%.

So although people worry that 30% of the budget goes to help the less fortunate, the real number is about 6%.

(And this is actually sort of depressing, when you think about it.)



Q: What do we really know about the budget?

What percentage of the federal budget in 2010 was spent on …?

	WHAT YOU THINK*	WHAT THE REAL NUMBERS ARE**
Military	30%	19.3%
Medicare	20%	13.1%
Social Security	20%	20.4%
Medicaid	15%	7.9%
Education	10%	2.7%
Foreign aid	10%	0.6%
Government pensions	10%	3.5%
Food assistance	10%	2.8%
Housing assistance	7%	1.7%
Public broadcasting	5%	0.01%

The majority of your taxes go to programs that benefit you and other middle-class Americans, such as Social Security and Medicare, and to programs that "benefit" you and other middle-class Americans, such as the military.

Part C: Political Issues

The Argument: *Government can't do anything right. Its forays into every field are tinged in failure. Whether it's trying to create contradictory "state owned businesses", funding pet projects that end up over budget and useless, or creating burdensome and ridiculous "consumer protection" rules, its heavy-handed actions are always detrimental and usually embarrassing.*

With this track record, what sane person would want to involve government in even more industries? The push to get government deeper into health care is a disaster waiting to happen, and could give us a chronically broken system like those in Europe, where people die because of bureaucratic inefficiency.

Other places from which we can profitably eliminate government's prying hands include our schools, our prisons, our gun dealerships, and the friendly neighborhood meth lab.

The Counterargument: *Government sometimes, though by no means always, does things right, and some of its institutions and programs are justifiably considered models of efficiency and human*

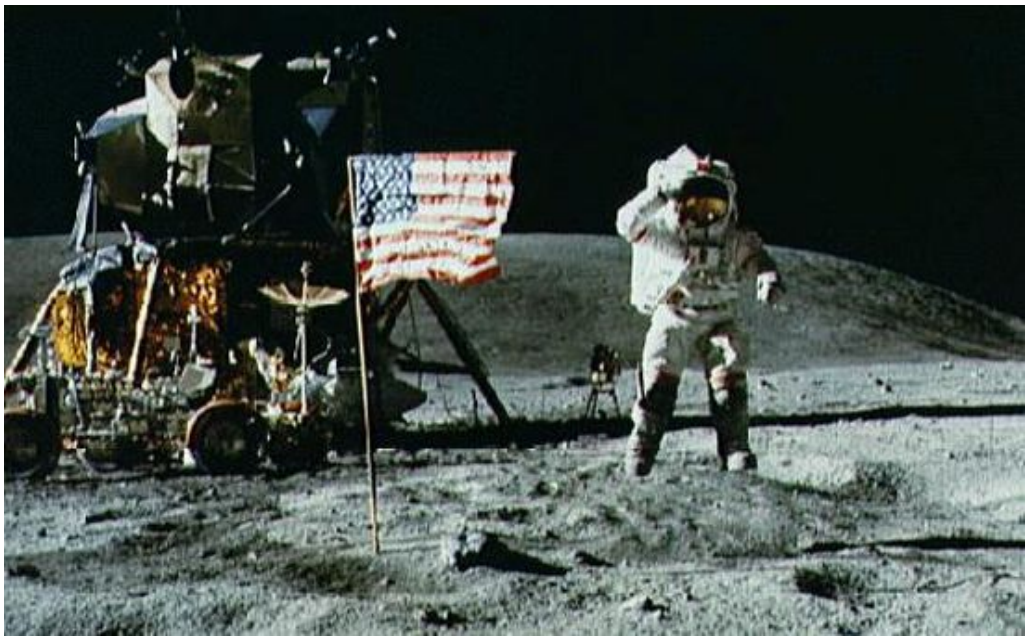
ingenuity. There are various reasons why people are less likely to notice these.

Government-run health systems empirically produce better health outcomes for less money than privately-run health systems for reasons that include economies of scale. There are a mountain of statistics that prove this. Although not every proposal to introduce government into health will necessarily be successful, we would do well to consider emulating more successful systems.

We should think twice about exactly how much government we are willing to remove from our schools, gun dealerships, and meth labs, and run away screaming at the proposal to privatize prisons.

7. Competence of Government

7.1: Government never does anything right.



7.1.1: Okay, fine. But that's a special case where, given an infinite budget, they were able to accomplish something that private industry had no incentive to try. And to their credit, they *did* pull it off, but do you have any examples of government succeeding at anything more *practical*?

Eradicating smallpox and polio globally, and cholera and malaria from their endemic areas in the US. Inventing the computer, mouse, digital camera, and email. Building the information superhighway

and the regular superhighway. Delivering clean, practically-free water and cheap on-the-grid electricity across an entire continent. Forcing integration and leading the struggle for civil rights. Setting up the Global Positioning System. Ensuring accurate disaster forecasts for hurricanes, volcanos, and tidal waves. Zero life-savings-destroying bank runs in eighty years. Inventing nuclear power *and* the game theory necessary to avoid destroying the world with it.

7.1.1.1: All right… all right… but apart from better sanitation and medicine and education and irrigation and public health and roads and a freshwater system and baths and public order… what has the government done for *us?*

Brought peace. But see also Government Success Stories and The Forgotten Achievements of Government

7.2: Large government projects are always late and over-budget.

The only study on the subject I could find, "What Causes Cost Overrun in Transport Infrastructure Projects?" (download study as .pdf) by Flyvbjerg, Holm, and Buhl, finds no difference in cost overruns between comparable government and private projects, and in fact find one of their two classes of government project (those not associated with a state-owned enterprise) to have a trend toward being *more* efficient than comparable private projects. They conclude that "…one conclusion is clear…the conventional wisdom, which holds that public ownership is problematic whereas private ownership is a main source of efficiency in curbing cost escalation, is dubious."

Further, when government cost overruns occur, they are not usually because of corrupt bureaucrats wasting the public's money. Rather, they're because politicians don't believe voters will approve their projects unless they spin them as being much cheaper and faster than the likely reality, leading a predictable and sometimes commendable execution to be condemned as "late and over budget" (download study as .pdf) While it is admittedly a problem that government provides an environment in which politicians have to lie to voters to

get a project built, the facts provide little justification for a narrative in which government is incompetent at construction projects.

7.3: State-run companies are always uncreative, unprofitable, and unpleasant to use.

Some of the greatest and most successful companies in the world are or have been state-run. Japan National Railways, which created the legendarily efficient bullet trains, and the BBC, which provides the most respected news coverage in the world as well as a host of popular shows like *Doctor Who*, both began as state-run corporations (JNR was later privatized).

In cases where state-run corporations are unprofitable, this is often not due to some negative effect of being state-run, but because the corporation was put under state control precisely because it was something so unprofitable no private company would touch it, but still important enough that it had to be done. For example, the US Post Office has a legal mandate to ship affordable mail in a timely fashion to every single god-forsaken town in the United States; obviously it will be out-competed by a private company that can focus on the easiest and most profitable routes, but this does not speak against it. Amtrak exists despite passenger rail travel in the United States being fundamentally unprofitable, but within its limitations it has done a relatively good job: on-time rates better than that of commercial airlines, 80% customer satisfaction rate, and double-digit year-on-year passenger growth every year for the past decade.

7.3.1: State-run companies may be able to paper-push with the best of them, but the government can never be truly innovative. Only the free market can do that. Look at Silicon Valley!

Advances invented either solely or partly by government institutions include, as mentioned before, the computer, mouse, Internet, digital camera, and email. Not to mention radar, the jet engine, satellites, fiber optics, artificial limbs, and nuclear energy. And that doesn't the less recognizable inventions used mostly in industry, or the scores of other inventions from government-funded universities and hospitals.

Even those inventions that come from corporations often come not from startups exposed to the free market, but from *de facto* state-owned monopolies. For example, during its fifty years as a state-sanctioned monopoly, the infamous Ma Bell invented (via its Bell Labs division) transistors, modern cryptography, solar cells, the laser, the C programming language, and mobile phones; when the monopoly was broken up, Bell Labs was sold off to Alcatel-Lucent, which after a few years announced it was cutting all funding for basic research to focus on more immediately profitable applications.

Although the media celebrates private companies like Apple as centers of innovation, Apple's expertise lies, at best, in consumer packaging. They did not invent the computer, the mp3 player, or the mobile phone, but they developed versions of these products that were attractive and easy to use. This is great and they deserve the acclaim and heaps of money they've gathered from their success, but let's make sure to call a spade a spade: they are good at marketing and design, not at brilliant invention of totally new technologies.

That sort of *de novo* invention seems to come mostly from very large organizations that can afford basic research without an obsession on short-term profitability. Although sometimes large companies like Ma Bell, invention-rich IBM and Xerox can fulfill this role, such organizations are disproportionately governments and state-sponsored companies, explaining their impressive track record in this area.

7.4: Most government programs are expensive failures.

I think this may be a form of media bias - not in the sense that some sinister figure in the media is going through and censoring all the stories that support one side, but in the sense that "Government Program Goes More Or Less As Planned" doesn't make headlines and so you never hear about it.

Let's say the government wants to spent $1 million to give food to poor children. If there are bureaucratic squabbles over where the money's supposed to come from, that's a headline. If they buy the food at above-market prices, that's a headline. If some corrupt

official manages to give the contract to provide the food to a campaign donor along the way, that's a *big* headline.

But what if none of these things happen, and poor children get a million dollars worth of food, and eat it, and it makes them healthier? I don't know about you, but I've never seen a headline about this. "Remember that time last year when Congress voted to give food to poor children. Well, they got it." What newspaper would ever publish something like that?

This is in addition to newspapers' desire to outrage people, their desire to sound "edgy" by pointing out the failures of the status quo rather than sounding like they're "pandering", and honestly that they're caught up in the same "government can never do anything right" narrative as everyone else.

Since every single time you ever hear about a government project it is always because that government project is going wrong, of course you feel like all government projects go wrong.

7.4.1: But a specific initiative to get money to the poor is one thing. What about a whole federal agency? We would know if it were failing, but we'd also be able to appreciate it when it succeeds, too.

Federal agencies that are successful sink into background noise, so that we don't think to thank them or celebrate them any more than we would celebrate that we have clean water (four billion people worldwide don't; thank the EPA and your local water board)

For example, the Federal Aviation Administration helps keep plane crashes at less than one per 21,000 years of flight time; you never think about this when you get on a plane. The National Crime Information Center collects and processes information about criminals from every police department in the country; you never think about this when you go out without being mugged. Zoning regulations, building codes, and the fire department all help prevent fires from starting and keep them limited when they do; you never think of this when you go the day without your house burning down.

One of government's major jobs is preventing things, and it's very hard to notice how many bad things *aren't* happening, until someone comes out with a report like e. coli poisoning has dropped by half in the past fifteen years. Even if you do hear the statistics, you may never think to connect them to the stricter food safety laws you wrote a letter to the editor opposing fifteen years ago.

7.4.2: You list cases where government regulation exists at the same time as a happy outcome, like the FAA and the lack of plane crashes, but that doesn't prove it was the regulation that *caused* the happy outcome.

No, it doesn't. For example, although workplace accidents have been cut in half since OSHA was founded, CATO wrote a very credible takedown in which they argue that was only a continuation of trends that have been going on since before OSHA existed.

Sometimes there are things we can do to identify cause. For example, as in the CATO study, we can compare trends before and after changes in government regulation; if there is a discontinuity, it may suggest the government was responsible. Second, we can compare trends in a country where a new regulation was introduced to trends in a country where it was not introduced; if the trend only changes in one country, that suggests an effect of the regulation. For example, after the FAA mandated "terrain awareness systems" in airplanes, the terrain-related accident rate sharply dropped to zero in the United States but was not affected in countries without similar rules.

But the important thing is that we apply our skepticism fairly and evenly: that we do not require mountains of evidence that a government regulation caused a positive result, while accepting that a regulation caused a negative result without a shred of proof.

It is very tempting for libertarians, when faced with anything going well even in a tightly regulated area, to say "Well, that just shows even this tight regulations can't hide how great private industry is!" and when anything goes wrong even in a very loosely regulated area, to say "Well, that just shows how awful regulation is, that even

a little of it can screw things up!" But this is unfair, and ignores that we do have some ways to disentangle cause and effect.

And in any case, there is still the difference between "Government destroys everything it touches" and "Everything government touches is doing pretty well, but you can't prove that it's directly caused by government action."

7.4.3: A lot of what government trumpets as "successful regulation" is just obvious stuff anyway that any individual in a free market would do of her own accord.

Very often, yesterday's regulation is today's obvious good idea that no one would dream of ignoring even if there were no regulation demanding it. But that neglects the role of government regulation in establishing social norms. Very often these are the regulations which those being regulated fought tooth and nail against at the time.

Many cars did not even *include* seatbelts until the government mandated that they do so. In 1983, the seat belt use rate in the United States was 14%. It was very clearly the government sponsored awareness campaigns and, later, mandatory seat belt laws that began being implemented around that era that raised seat belt rates; we know because we can watch the statistics state in different states as their legislation either led the campaign or lagged behind it.

After almost three decades of intense government pressure on automakers to allow and promote seatbelts, and on motorists to use them, seatbelt rates are now as high as 85%.

According to estimates, seatbelts save about 11,000 lives a year in the US. Different studies estimate between 80,000 and 100,000 lives saved in the last decade alone. For some perspective that's the number of American deaths from 9/11 + the Vietnam War + both Iraq Wars + the Afghanistan War + Hurricane Katrina.

I completely acknowledge that if the government completely dropped all seatbelt regulations tomorrow, automakers would continue putting seatbelts in cars, and drivers would keep wearing them. That doesn't mean government is useless, that means

government, the only entity big enough to effect a nationwide change not just in behaviors but in social norms, did its job very very well.

8. Health Care

8.1: Government would do a *terrible* job in health care. We should avoid government-run "socialized" medicine unless we want cost overruns, long waiting times, and death panels.

Government-run health systems empirically do better than private health systems, while also costing much less money.

Let's compare, for example, Sweden, France, Canada, the United Kingdom, and the United States. The first four all have single-payer health care (a version of government-run health system); the last has a mostly private health system (although it shouldn't matter, we'll use statistics from before Obamacare took effect). We'll look at three representative statistics commonly used to measure quality of health care: infant mortality, life expectancy, anc cancer death rate.

Infant mortality is the percent of babies who die in the first few weeks of life, usually a good measure of pediatric and neonatal care. Of the five countries, Sweden has the lowest infant mortality at 2.56 per 1,000 births, followed by France at 3.54, followed by the UK at 4.91, followed by Canada at 5.22, with the United States last at 6.81. (source)

Life expectancy, the average age a person born today can expect to live, is a good measurement of lifelong and geriatric care. Here Sweden is again first at 80.9, France and Canada tied for second at 80.7, the UK next at 79.4, and the United States once again last at 78.3. (source)

Taking cancer deaths per 100,000 people per year as representative of deaths from serious disease, here we find the UK doing best at 253.5 deaths, Sweden second at 268.2, France in third at 286.1, and the United States again in last place at 321.9 deaths (source: OECD statistics; data for Canada not available).

So we notice that the United States does worse than all four countries with single-payer health systems, even though America is wealthier per capita than any of them. This is not statistical cherry-picking: any way you look at it, the United States has one of the least effective health systems in the developed world.

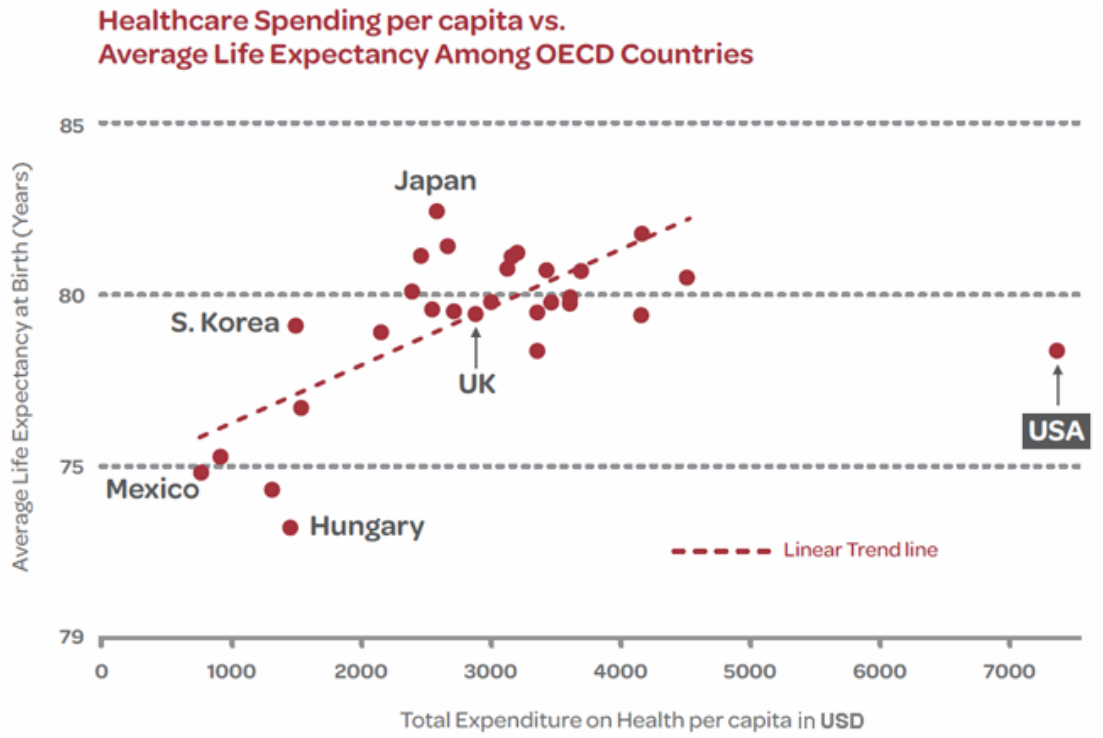
8.2: Government-run health care would be bloated, bureaucratic, and unnecessarily expensive, as opposed to the sleek, efficient service we get from the free market.

Actually, government-run health care is empirically more efficient than market health care. For example, Blue Cross New England employs more people to administer health insurance for its 2.5 million customers than the Canadian health system employs to administer health insurance for 27 million Canadians. Health care spending per person (public + private) in Canada is half what it is in America, yet Canadians have longer life expectancy, lower infant mortality, and are healthier by every objective standard.

Remember those five countries from the last question?

The UK spends $1,675 per person per year on health care. Canada spends $1,939. Sweden, which you'll remember did best on most of the statistics, spends $2,125. France spends $2,288. Americans spend on average $4,271 - almost *three times* as much as Britain, a country which *delivers better health care*.

When this argument gets put in graph form, it becomes even clearer that US health inefficiency is literally off the chart.



If these were companies in the free market, the company that charges three times as much to provide a worse service would have gone bankrupt long ago. That company is American-style private health care.

8.3: In government-run health care, people are relegated to "waiting lists", where they have to wait months or even years for doctor visits, surgeries, and other procedures. Sometimes people die on these waiting lists. Obviously, this is unacceptable and a knock-down argument against government-run health care.

The laws of supply and demand apply in health care as much as anywhere else: people would like to see doctors as quickly as possible, but doctors are a scarce resource that must be allocated somehow.

In a private system, doctor access is allocated based on money; this has the advantage of incentivizing the production of more doctors and of ensuring that people with enough money can see doctors quickly. These are also its disadvantages: assuming more people want to see a doctor than need to do so, costs will spiral out of control and poor people will have limited or no access.

In a public system, doctor access is allocated based on medical need. Although no one will be turned away from a doctor in an emergency situation, people may have to wait a long amount of time for elective surgeries in order that other sicker people, including poor people who would not be seen at all in a private system, can be seen first.

The relative effectiveness of the two systems can once again be seen in the infant mortality, life expectancy, and cancer survival rate statistics.

8.4: Government-run health care inevitably includes "death panels" who kill off expensive patients in order to save money on health care costs.

The private system as it exists now in America also has bodies that make these kinds of rationing decisions. Health care rationing is not some sinister conspiracy but a reasonable response to limited resources. The complete argument is here, but I can sum up the basics:

Insurance providers, whether they are a government agency or a private corporation, have a finite amount of money; they can only spend money they have. In one insurance company, customers might pay hundred million dollars in fees each year, so the total amount of money the insurance company can spend on all its customers that year is a hundred million dollars. In reality, since it is a business, it wants to make a profit. Let's say it wants a profit of ten percent. That means the total amount of money it has to spend is ninety million dollars.

But as a simplified example, let's reduce this to an insurance company with one hundred customers, each of whom pays $1. This insurance company wants 10% profit, so it has $90 to spend (instead of our real company's $90 million). Seven people on the company's plan are sick, with seven different diseases, each of which is fatal. Each disease has a cure. The cures cost, in order, $90, $50, $40, $20, $15, $10, and $5.

We are far too nice to ration health care with death panels; therefore, we have decided to give everyone every possible treatment. So when the first person, the one with the $90 disease, comes to us, we gladly spend $90 on their treatment; it would be inhuman to just turn them away. Now we have no money left for anyone else. Six out of seven people die.

The fault here isn't with the insurance company wanting to make a profit. Even if the insurance company gave up its ten percent profit, it would only have $10 more; enough to save the person with the $10 disease, but five out of seven would still die.

A better tactic would be to turn down the person with the $90 disease. Instead, treat the people with $5, $10, $15, $20, and $40 diseases. You still use only $90, but only two out of seven die. By refusing treatment to the $90 case, you save four lives. This solution can be described as more cost-effective; by spending the same amount of money, you save more people. Even though "cost-effectiveness" is derided in the media as being opposed to the goal of saving lives, it's actually all about saving lives.

If you don't know how many people will get sick next year with what diseases, but you assume it will be pretty close to the amount of people who get sick this year, you might make a rule for next year: Treat everyone with diseases that cost $40 or less, but refuse treatment to anyone with diseases that cost $50 or more.

This rule remains true in the case of the $90 million insurance company. In their case, no one patient can use up all the money, but they still run the risk of spending money in a way that is not cost-effective, causing many people to die. Like the small insurance company, they can increase cost-effectiveness by creating a rule that they won't treat people with diseases that cost more than a certain amount.

So, as one commentator pointed out, "death panels" should be called "life panels": they aim to maximize the total number of lives that can be saved with a certain limited amount of resources.

8.5: Why is government-run health care so much more effective?.

A lot of it is economies of scale: if the government is ensuring the entire population of a country, it can get much better deals than a couple of small insurance companies. But a lot of it is more complicated, and involves people's status as irrational consumers of health products. A person sick with cancer doesn't want to hear a cost-benefit analysis suggesting that the latest cancer treatment is probably not effective. He wants that treatment right now, and the most successful insurance companies and hospitals are the ones that will give it to him. Here's a good article explaining some of the systematic flaws in the economics of health care under the American system.

It could also be that really good health care and the profit motive don't mix: studies show that for-profit hospitals are more expensive, and have poorer care (as measured in death rates) than not-for-profit hospitals.

9. Prison Privatization

9.1: Privatized, for-profit prisons would be a great way to save money.

No one likes criminals very much. Even so, most of us agree that even criminals deserve humane conditions. We reject cruel and unusual punishment, and try to keep prisoners relatively warm, clean, and well-fed. This is not only a moral issue, but a practical one: we don't want prisoners to go insane or suffer breakdowns, because we want them to be able to re-adjust into normal society after they are released.

For-profit prisons have all of the flaws of for-profit companies with none of the advantages. Normal companies want to cut costs wherever possible, but this is balanced by customer satisfaction: if they treat their customers poorly or create a low-quality product, they won't make money. In prisons, the ability to get new "customers" comes completely uncoupled from the quality of the

product they provide. If the government pays them a certain fixed amount per prisoner, the prison's only way to increase profits is by treating prisoners as shabbily as possible without killing them. Indeed, statistics show that prisoners in private prisons have worse medical care, terrible living conditions, and rates of in-prison violence 150% greater than those in public prisons. Private prisons refuse to collect data on recidivism rates, but a moment's thought reveals that they have an economic incentive to keep them as *high* as possible.

But the real dangers lie in the corruptibility of the political process, something with which libertarians are already familiar. Private prisons have been active in lobbying for stricter sentencing guidelines like the Three Strikes Law, which encourages governments to imprison criminals for life. In a country that already imprisons more of its population than any other country in the world, it is extremely dangerous to create a powerful political force whose self-interest lies in imprisoning as many people as possible.

But the most striking example of the danger of private prisons is the case of two judges who received bribes from private prisons to jail innocent people.

If this is the alternative, I'm willing to bite the bullet and accept the overpaid prison guards with annoying unions who dominate the public prisons.

9.2: What? Libertarians don't actually believe in private prisons!

Fair enough; I got this complaint a few times on the first version and I acknowledge it's not an integral component of libertarian philosophy. I included it because it seems to stem from the same "government can never do anything right and we should privatize everything" idea that drives a lot of libertarian thinking, and because I really, really don't like private prisons.

10. Gun Control

10.1: Gun control laws only help criminals, who are not known for following laws in any case, make sure that their victims are unarmed and unable to resist; as such, they increase crime.

The statistics supporting this view seem relatively solid and I agree that attempts to ban or restrict access to guns are a bad idea.

On the other hand, many of the issues surrounding gun control are much less restrictive. For example, some involve restrictions on sales to criminals, "cooldown periods" before purchase, mandatory safety training, et cetera.

Although I haven't seen any evidence either way on whether these laws are beneficial, they should be evaluated on their own merits rather than as part of a narrative in which all gun laws must be opposed because gun control is bad.

11. Education

11.1: Government sponsored public education is a horrible failure.

Compared to what?

Compared to the period when there *wasn't* government-sponsored public education…well, that's hard to say because of poor statistic-keeping at that time, and how one counts minorities and women, who usually weren't educated at all back then. The most official statistics (eg NOT the ones you find without citation on libertarian blogs that say literacy was 100% way back when and became abysmal as soon as public schooling started) say that white illiteracy declined from about 11.5% in the mid-1800s to about 0.5% in 1980, and black illiteracy from about 80% to 1.5% over the same period.

Compared to other countries, the US does relatively poorly considering its wealth, but all the other countries that do better than the US *also* have government-sponsored public education, sometimes to a much greater degree than we do.

Compared to private schools, public schools actually do better once confounders like race, class, and income have been adjusted out of

the analysis.

(Yes, without such adjustment private schools do better - but considering that private schools cater towards wealthy students - who usually do better in school - and often have selective admission policies in which they only take students who are already pretty smart - whereas public schools have to take everyone including dumb kids, kids with learning disabilities, and kids from broken families in ghettos - such unadjusted data is meaningless. It's the equivalent of noting that the doctor who specializes in acne has fewer patients die than the doctor who specializes in cancer: it's not that she's a better doctor, just that she only takes cases who are pretty healthy already.)

Our educational system certainly has immense room for improvement. But the country that consistently tops world education rankings, Finland, has zero private schools (even all the universities are public) and no "school choice". What it does have is extremely well-credentialed, highly paid teachers (and, unfortunately, an ethnically homogenous population without any dire poverty or broken families, which probably counts for a heck of a lot more than anything else). So whatever America's specific failures or successes, the mere existence of public education is not a credible scapegoat.

11.2: Why not dismantle the public education system and have a voucher system that offers parents free choice over where to send their kids?

I think this idea has merit, and that we should at least experiment with it and see if it works. That having been said, I do see one huge caveat.

Libertarians tend not to believe in *equality of results* - they think it's okay if more skilled people are more successful - but one of the qualities I most admire about them is that they usually do believe in equality of opportunity: that everyone gets an equal chance at life. I mentioned before how inheriting money from your parents can complicate that, but it would be ethically complicated to try and

"solve" that problem, so it might be the sort of thing we just have to live with.

But imagine if your parents chose where to send you for school. Even if we somehow eliminated the cost issue by making everyone accept a school voucher of equal value, clever parents would compare the pros and cons of various schools and send their child to the best one. Not-so-clever parents would get fooled by TV commercials with sexy celebrities and send their kids to terrible schools. Super religious parents would send their kids to schools that taught only religious education and shunned math and science and history as the evil trappings of the secular world. Muslim parents would send their kids to madrassas. Immigrant parents might send their kids to Spanish-only schools so that they didn't drift too far away from their families. Parents with strong political beliefs could send their kids to schools that did their best to brainwash their kids into having the same beliefs as them.

And there *would* be kids who succeeded in spite of all this, who made it through twelve years of constant brainwashing and ignorance, and somehow managed to become intelligent adults who could learn all the education they missed during their free time. But statistically, there wouldn't be very many of them, any more than there were a bunch of Christians in Saudi Arabia in the example a few pages back.

Right now, parents can screw up lots of facets of a kid's life, but they can only do so much to screw up their education. And I have this vague hope that maybe a kid with horrible parents, if she was exposed to decent people and a free exchange of ideas in school might be able to use that brief period of respite to gain a foothold on sanity.

So what I'm saying is, if there were school choice, if we wanted to protect equality of opportunity and childrens' rights, we'd probably have to regulate the heck out of them, which to some degree would defeat the point.

11.3: I don't believe the government should be in the business of "protecting" children from their parents.

You should. It's a pretty important business, even if you subscribe to libertarian assumptions. Even libertarians tend to agree that the government should generally be protecting people from slavery and from the use of force.

Children are basically slaves to their parents for the first ten to fifteen years of their lives, and parents have a special social permission to use force against their children.

In the best possible case, this is an incredibly silly metaphor and one no one would ever even think about. In the worst possible case, it's completely and literally true.

I have met people with horrible parents. The first eighteen years (or less, if they were able to get themselves legally emancipated early) of their lives were a living hell. These are people who literally have control of every single thing you do, from whether you can eat dinner to who you are allowed to make friends with to what church you go to to what opinions you can express to whether you're allowed to sleep at night. They are people who can torture and beat you to within an inch of your life, and maybe a social worker will take you away for a few months, and then that social worker will probably return you right back to them. And if it's just *emotional* torture, you can forget about even getting the social worker. Here I am writing a FAQ called "Why I Hate Your Freedom", and even *I* shudder to think about this.

And obviously the parent-child relationship is a healthy one in 99% of cases, and child-rearing has been around since deep prehistoric time, and we would be idiots to mess with it, and no one wants a dystopia where the government takes kids from their parents and raises them in a commune or whatever.

But unless you think rights and morality only start existing on someone's eighteenth birthday, if if there were *one* form of government intervention that even libertarians should be able to get

behind, it would be protecting children from their parents, in the rare few cases where this is necessary.

Part D: Moral Issues

The Argument: *Moral actions are those which do not initiate force and which respect people's natural rights. Government is entirely on force, making it fundamentally immoral. Taxation is essentially theft, and dictating the conditions under which people may work (or not work) via regulation is essentially slavery. Many government programs violate people's rights, especially their right to property, and so should be opposed as fundamentally immoral regardless of whether or not they "work".*

The Counterargument: *Moral systems based only on avoiding force and respecting rights are incomplete, inelegant, counterintuitive, and usually riddled with logical fallacies. A more sophisticated moral system, consequentialism, generates the principles of natural rights and non-initiation of violence as heuristics that can be used to solve coordination problems, but also details under what situations such heuristics no longer apply. Many cases of government intervention are such situations, and so may be moral.*

12. Moral Systems

12.1: Freedom is incredibly important to human happiness, a precondition for human virtue, and a value almost everyone holds dear. People who have it die to protect it, and people who don't have it cross oceans or lead revolutions in order to gain it. But government policies all infringe upon freedom. How can you possibly support this?

Freedom is one good among many, albeit an especially important one.

In addition to freedom, we value things like happiness, health, prosperity, friends, family, love, knowledge, art, and justice. Sometimes we have to trade off one of these goods against another. For example, a witness who has seen her brother commit a crime

may have to decide between family and justice when deciding whether to testify. A student who likes both music and biology may have to decide between art and knowledge when choosing a career. A food-lover who becomes overweight may have to decide between happiness and health when deciding whether to start a diet.

People sometimes act as if there is some hierarchy to these goods, such that Good A always trumps Good B. But in practice people don't act this way. For example, someone might say "Friendship is worth more than any amount of money to me." But she might continue working a job to gain money, instead of quitting in order to spend more time with her friends. And if you offered her $10 million to miss a friend's birthday party, it's a rare person indeed who would say no.

In reality, people value these goods the same way they value every good in a market economy: in comparison with other goods. If you get the option to spend more time with your friends at the cost of some amount of money, you'll either take it or leave it. We can then work backward from your choice to determine how much you *really* value friendship relative to money. Just as we can learn how much you value steel by learning how many tons of steel we can trade for how many barrels of oil, how many heads of cabbages, or (most commonly) how many dollars, so we can learn how much you value friendship by seeing when you prefer it to opportunities to make money, or see great works of art, or stay healthy, or become famous.

Freedom is a good much like these other goods. Because it is so important to human happiness and virtue, we can expect people to value it very highly.

But they do not value it infinitely highly. Anyone who valued freedom from government regulation infinitely highly would move to whichever state has the most lax regulations (Montana? New Hampshire?), or go live on a platform in the middle of the ocean where there is no government, or donate literally all their money to libertarian charities or candidates on the tiny chance that it would effect a change.

Most people do not do so, and we understand why. People do not move to Montana because they value aspects of their life in non-Montana places - like their friends and families and nice high paying jobs and not getting eaten by bears - more than they value the small amount of extra freedom they could gain in Montana. Most people do not live on a platform in the middle of the ocean because they value aspects of living on land - like being around other people and being safe - more than they value the rather large amount of extra freedom the platform would give them. And most people do not donate literally all their money to libertarian charities because they like having money for other things.

So we value freedom a finite amount. There are trade-offs of a certain amount of freedom for a certain amount of other goods that we already accept. It may be that there are other such trade-offs we would also accept, if we were offered them.

For example, suppose the government is considering a regulation to ban dumping mercury into the local river. This is a trade-off: I lose a certain amount of freedom in exchange for a certain amount of health. In particular, I lose the freedom to dump mercury into the river in exchange for the health benefits of not drinking poisoned water.

But I don't really care that much about the freedom to dump mercury into the river, and I care a lot about the health benefits of not drinking poisoned water. So this seems like a pretty good trade-off.

And this generalizes to an answer to the original question. I completely agree freedom is an extremely important good, maybe the most important. I don't agree it's an infinitely important good, so I'm willing to consider trade-offs that sacrifice a small amount of freedom for a large amount of something else I consider valuable. Even the simplest laws, like laws against stealing, are of this nature (I trade my "freedom" to steal, which I don't care much about, in exchange for all the advantages of an economic system based on private property).

The arguments above are all attempts to show that some of the trade-offs proposed in modern politics are worthwhile: they give us enough other goods to justify losing a relatively insignificant "freedom" like the freedom to dump mercury into the river.

12.1.1: But didn't Benjamin Franklin say that those who would trade freedom for security deserve neither?

No, he said that those who would trade *essential* liberty for *temporary* security deserved neither. Dumping mercury into the river hardly seems like essential liberty. And when Franklin was at the Constitutional Convention he agreed to replace the minimal government of the Articles of Confederation with a much stronger centralized government just like everyone else.

12.2: Taxation is theft. And when the government forces you to work under their rules, for the amount of money they say you can earn, that's slavery. Surely you're not in favor of theft and slavery.

Consider the argument "How can we have a holiday celebrating Martin Luther King? After all, he was a *criminal*!"

Technically, Martin Luther King *was* a criminal, in that he broke some laws against public protests that the racist South had quickly enacted to get rid of him. It's why he famously spent time in Birmingham Jail.

And although "criminal" is a very negative-sounding and emotionally charged word, in this case we have to step back from our immediate emotional reaction and notice that the ways in which Martin Luther King was a criminal don't make him a worse person.

A philosopher might say we're equivocating between two meanings of "criminal", one meaning of "person who breaks the law", and another meaning of "horrible evil person." Just because King satisfies the first meaning (he broke the law) doesn't mean he has to satisfy the second (be horrible and evil).

Or consider the similar argument: "Ayn Rand fled the totalitarian Soviet Union to look for freedom in America. That makes her a

traitor!" Should we go around shouting at Objectivists "How can you admire Ayn Rand when she was a dirty rotten *traitor*"?

No. Once again, although "traitor" normally has an automatic negative connotation, we should avoid instantly judging things by the words we can apply to them, and start looking at whether the negative feelings are deserved.

Or once again the philosopher would say we should avoid equivocating between "traitor" meaning "someone who switches sides from one country to an opposing country" and "horrible evil untrustworthy person."

Our language contains a lot of words like these which package a description with a moral judgment. For example, "murderer" (think of pacifists screaming it at soldiers, who do fit the technical definition "someone who kills someone else"), "greedy" (all corporations are "greedy" if you mean they would very much like to have more money, but politicians talking about "greedy corporations" manage to transform it into something else entirely) and of course that old stand-by "infidel", which sounds like sufficient reason to hate a member of another religion, when in fact it simply means a member of another religion. It's a stupid, cheap trick unworthy of anyone interested in serious rational discussion.

And calling taxation "theft" is exactly the same sort of trick. What's theft? It's taking something without permission. So it's true that taxation is theft, but if you just mean it involves taking without permission, then everyone from Lew Rockwell up to the head of the IRS already accepts that as a given.

This only sounds like an argument because the person who uses it is hoping people will let their automatic negative reaction to theft override their emotions, hoping they will equivocate from theft as "taking without permission" to "theft as a terrible act worthy only of criminals".

Real arguments aren't about what words you can apply to things and how nasty they sound, real arguments about what good or bad

consequences those things produce.

12.3: Government actions tend to involve the initiation of force against innocent people. Isn't that morally wrong?

Why should it be morally wrong?

12.3.1: Because the initiation of force always has bad consequences, like ruining the economy or making people unhappy.

Sometimes it does. Other times it has good consequences.

Take cases like the fish farming, boycott, and charity scenarios above. There the use of force to solve the coordination problem meets an extraordinarily strict set of criteria: not only does it benefit the group as a whole, not only does it benefit every single individual in the group, but every single individual in the group knows that it benefits them and endorses that benefit (eg would vote for it).

In other cases, such as the retirement savings example above, the use of force meets only a less strict set of criteria: it benefits the group as a whole, it benefits every single individual in the group, but not every individual in the group necessarily knows that it benefits them or endorses that benefit. These are the cases libertarians might call "paternalism".

Still more cases satisfy an even looser criterion. They benefit the group as a whole, but they might not benefit every single individual in the group, and might harm some of them. These are the cases that libertarians might call "robbing Peter to pay Paul".

All three of these sets of cases belie the idea that the use of force must on net have bad consequences.

12.3.2: Okay, maybe it's wrong because some moral theory that's not about consequences tells me it's wrong.

If your moral theory doesn't involve any consequences, why follow it? It seems sort of like an arbitrary collection of rules you like.

The Jews believe that God has commanded them not to murder. They also believe God has commanded them not to start fires on Saturdays. Jews who lose their belief in God usually continue not to murder, but stop worrying about whether or not they light fires on Saturdays. Likewise, evangelical Christians believe stealing is a sin, and that homosexuality is also a sin. If they de-convert and become atheists, most of them will still oppose stealing, but most will stop worrying about homosexuality. Why?

Killing and stealing both have bad consequences; in fact, that seems to be the essence of why they're wrong. Fires on Saturday and homosexuality don't hurt anybody else, but killing and stealing do.

Why are consequences to other people seems such a specially relevant category? The argument is actually itself pretty libertarian. I can do whatever I want with my own life, which includes following religious or personal taboos. Other people can do whatever they want with their own lives too. The stuff that matters - the stuff where we have to draw a line in the sand and say "Nope, this is moral and this is immoral, doesn't matter what you think" is because it has some consequence in the real world like hurting other people.

12.3.2.1: I was always taught that the essence of morality was the Principle of Non-Aggression: no one should ever initiate force, except in self-defense. What exactly is wrong with this theory?

At least two things. First, once you disentangle it from the respect it gets as the Traditional Culturally Approved Ground Of Morality, the actual rational arguments for it as a principle are surprisingly weak. Second, in order to do anything practical with it you need such a mass of exceptions and counter-exceptions and stretches that one starts to wonder whether it's doing any philosophical work at all; it becomes a convenient hook upon which to hang our pre-existing prejudices rather than a useful principle for solving novel moral dilemmas.

12.3.2.1.1: What do you mean by saying that the rational arguments for the Principle of Non-Aggression are weak?

There are dozens of slightly different versions of these arguments, and I don't want to get into all of them here, so I'll concentrate on the most common.

Some people try to derive the Principle of Non-Aggression from self-ownership. But this is circular reasoning: the form of "private property" you need to own anything, including your self/body, is a very complicated concept and one that requires some form of morality in order to justify; you can't use your idea of private property as a justification for morality. Although it's obvious that in some sense you *are* your body, there's no way to go from here to "And therefore the proper philosophical relationship between you and your body is the concept of property exactly as it existed in the 17th century British legal system."

This also falls afoul of the famous is-ought dichotomy, the insight that just because something *is* true doesn't mean it *should be* true. Just because we notice some factual relationship between yourself and your body doesn't mean that relationship between yourself and your body is good or important or needs to be protected in laws. We might eventually *decide* it should be (and hopefully we will!) but we need to have other values in order to come to that decision; we can't use the decision as a *basis* for our values.

The self-ownership argument then goes from this questionable assumption to other even more questionable ones. If you use your body to pick fruit, that fruit becomes yours, even though you didn't make it. If you use your body to land on Tristan de Cunha and plant a flag there and maybe pick some coconuts, that makes Tristan de Cunha and everything on your property and that of your heirs forever, even though you *definitely* didn't make the island. And if someone else lands on Tristan de Cunha the day after you, you by right control every facet of their life on the island and they have to do whatever you say or else leave. There are good arguments for why some of these things make economic sense, but they're all practical arguments, not moral ones positing a necessary relationship.

Oddly enough, although apparently your having a body does license you to declare yourself Duke of Tristan de Cunha, it doesn't license you to use your fist to punch your enemy in the gut, or use your legs to walk across a forest someone else has said they claim, even though your ability to move your hand rapidly in the direction of your enemy's abdomen, or your feet along a forest path, seems like a much more fundamental application of your body than taking over an island.

All of these rules about claiming islands and not punching people you don't like and so on are potentially good rules, but trying to derive them just from the fact that you have a body starts to seem a bit hokey.

12.3.2.1.2: What do you mean by saying that the Non-Aggression Principle requires so many exceptions and counter-exceptions that it becomes useless except as a hook upon which to hang prejudices we from other sources?

First, the principle only even slightly makes sense by defining "force" in a weird way. The NAP's definition of "force" includes walking into your neighbor's unlocked garden when your neighbor isn't home and picking one of her apples. It includes signing a contract promising to deliver a barrel of potatoes, but then not delivering the potatoes when the time comes. Once again, I agree these are bad things that we need rules against. But it takes quite an imagination to classify them under "force", or as deriving from the fact that you have a body. This is a good start to explaining what I mean when I say that people *claim* that they're using the very simple-sounding "no initiation of force" principle but are actually following a more complicated and less justified "no things that seem bad to me even though I can't explain why".

Second, even most libertarians agree it can be moral to initiate force in certain settings. For example, if the country is under threat from a foreign invader or from internal criminals, most libertarians agree that it is moral to levy a small amount of taxation to support an army or police force that restores order. Again, this is a very good idea -

but also a blatant violation of the Non-Aggression Principle. When libertarians accept the initiation of force to levy taxes for the police, but protest that initiating force is always wrong when someone tries to levy taxes for welfare programs, it reinforces my worry that the Non-Aggression Principle is something people *claim* to follow while actually following their own "no things that seem bad to me even though I can't explain why, but things that seem good to me are okay" principle.

(I acknowledge that some libertarians take a stand against taxes for the military and the police. I admire their consistency even while I think their proposed policies would be a disaster.)

Third, when push comes to shove the Non-Aggression Principle just isn't strong enough to solve hard problems. It usually results in a bunch of people claiming conflicting rights and judges just having to go with whatever seems intuitively best to them.

For example, a person has the right to live where he or she wants, because he or she has "a right to personal self-determination". Unless that person is a child, in which case the child has to live where his or her parents say, because…um…the parents have "a right to their child" that trumps the child's "right to personal self-determination". But what if the parents are evil and abusive and lock the child in a fetid closet with no food for two weeks? Then maybe the authorities can take the child away because…um…the child's "right to decent conditions" trumps the parents' "right to their child" even though the latter trumps the child's "right to personal self-determination"? Or maybe they can't, because there shouldn't even be authorities of that sort? Hard to tell.

Another example. I can build an ugly shed on my property, because I have a "right to control my property", even though the sight of the shed leaves my property and irritates my neighbor; my neighbor has no "right not to be irritated". Maybe I can build a ten million decibel noise-making machine on my property, but maybe not, because the noise will leave my property and disturbs neighbor; my "right to control my property" might or might not trump my neighbor's "right

not to be disturbed", *even though disturbed and irritated are synonyms*. I definitely can't detonate a nuclear warhead on my property, because the blast wave will leave my property and incinerates my neighbor, and my neighbor apparently does have a "right not to be incinerated".

If you've ever seen people working within our current moral system trying to solve issues like these, you quickly realize that not only are they making it up as they go along based on a series of ad hoc rules, but they're so used to doing so that they no longer realize that this is undesirable or a shoddy way to handle ethics.

12.4: Is there a better option than the Non-Aggression Principle?

Yes. It's consequentialism, the principle that it is moral to do whatever has, on net, the best consequences. This is about equivalent to saying "to do whatever makes the world a better place". It's the principle we've been using implicitly throughout this FAQ and the principle most people use implicitly throughout their lives.

It's also the principle that drives capitalism, where people are able to create incredible businesses and innovations because they are trying to do whatever has the best financial consequences for themselves. Consequentialism just takes that insight and says that instead of just doing it with money, let's do it with everything we value.

12.4.1: Best consequences according to whom?

Well, if you're the one making the moral decision, then best consequences according to you. All it's saying is that your morality should be a reflection of your value system and your belief in a better world. Your job as a moral agent is to try to make the world a better place by whatever your definition of "better place" might be.

Sticking to the capitalism analogy, consumerism "tells you" (not that you need to be told) to get whatever goods you value most. Consequentialism does the same, but tells you to try to get the collection of abstract moral goods you value the most.

But remember our discussion of trade-offs above. Most people value many different moral goods, and you are no exception. If you're

trying to make the world a better place, you should be thinking about your relative valuation of all these goods and what trade-offs you are willing to make.

12.4.2: Best consequences for me, or best consequences for everyone?

Again, this is your decision. If you're completely selfish, then consequentialism tells you to seek out the best consequences for yourself. This probably wouldn't mean being a libertarian - thankless activism for an unpopular political position is really a *terrible* way to go about looking out for Number One. It would probably mean cheating off the government - either in the form of welfare abuse if you're poor and lazy, or in the form of crony capitalism if you're rich and ambitious. As icing on the cake, make sure to become a sanctimonious and hypocritical liberal, as it's a great way to become popular and get invited to all the fancy parties.

But if you care about people other than yourself, consequentialism tells you to seek out the best consequences for the people you care about (which could be anything from your family to your country to the world). This could involve political activism, and it could even involve political activism in favor of libertarianism if you think it's the best system of government.

Alternately, it could justify trying to *start* a government, if there's no government yet and you think a world with government would be better for the people you care about than one without it.

Most of the rest of this section will be assuming you do in fact care for other people at least a little.

12.4.3: Since many people probably want different things and care about different people, don't we end out in a huge war of all against all until either everyone is dead or one guy is dictator?

Would that be a good consequence? If not, people who try to promote good consequences and make the world a better place would try to avoid it.

Because this world of violence and competition is so obviously a bad consequence, any consequentialist who gives it a moment's thought agrees not to start a huge war of all against all that ends with everyone dead or one guy as dictator by binding themselves by moral rules whenever binding themselves by those moral rules seems like it would have good consequences or make the world a better place; see Section 13 for more.

12.4.4: Doesn't that sound a lot like "the ends justify the means"? Wouldn't it lead to decadence, slavery, or some other dystopia?

Once again, if you consider dictatorship, slavery, and dystopia to be bad consequences, then by definition following this rule is the best way to *avoid* doing that.

The rule isn't "do whatever sounds like it would have the best consequences if you have an IQ of 20 and refuse to think about it for even five seconds", it's "do what would *actually* have the best consequences. Sometimes this involves admitting human ignorance and fallibility and *not* pursuing every hare-brained idea that comes into your head.

12.4.5: Okay, okay, I understand that if people did what *actually* had good consequences it would have good consequences, but I worry that if people do what they *think* has good consequences, it will lead to violence and dictatorship and dystopia and all those other things you mentioned above.

Yes, I agree this is an important distinction. There are two uses for a moral system. The first is to define what morality is. The second is to give people a useful tool for choosing what to do in moral dilemmas. I am arguing that consequentialism does the first. I don't think it does the second right out of the box.

To try a metaphor, doctors sometimes have two ways of defining disease; the gold standard and the clinical standard. The gold standard is the "perfect" test for the disease; for example, in Alzheimers disease, it's to autopsy the brain after the person has

died and see if it has certain features under the microscope. Obviously you can't autopsy a person who's still alive, so when doctors are actually trying to diagnose Alzheimers they use a more practical method, like how well the person does on a memory test.

Right now I'm arguing that consequentialism is the gold standard for morality: it's the purest, most sophisticated explanation of what morality actually is. At the same time, it might be a terrible idea to make your everyday decisions based on it, just as it's a terrible idea to diagnose Alzheimers with an autopsy in someone who's still alive.

However, once we know that consequentialism is the gold standard for morality, we can start designing our clinical standards by trying to figure out which "clinical standard" for morality will produce the best consequences. See Section 13 for more.

12.4.6: I still am not completely on board with consequentialism, or I'm not sure I understand it.

For more information on consequentialism, see the sister document to this FAQ, the Consequentialism FAQ.

13. Rights and Heuristics

13.1: Is there a moral justification for rights, like the right to free speech or the right to property?

Yes. Rights are the "clinical standard" for morality, the one we use to make our everyday decisions after we acknowledge that pure consequentialism might not lead to the best consequences when used by fallible humans.

In this conception, rights are conclusions rather than premises. They are heuristics (heuristic = a rule-of-thumb that usually but not always works) for remembering what sorts of things usually have good or bad consequences, a distillation of moral wisdom that is often more trustworthy than morally fallible humans.

For example, trying to tell people what religions they can or can't follow almost always has bad consequences. At best, people are

miserable because they're being forced to follow a faith they don't believe in. At worst, they resist and then you get Inquisitions and Holy Wars and everyone ends up dead. Restriction of religion causing bad consequences is sufficiently predictable that we generalize it into a hard and fast rule, and call that rule something like the "right to freedom of religion".

Other things like banning criticism of the government, trying to prevent people from owning guns, and seizing people's property willy-nilly also work like this, so we call those "rights" too.

13.2: So if you think that violating rights will have good consequences, then it's totally okay, right?

It's not quite so simple. Rights are not just codifications of the insight that certain actions lead to bad consequences, they're codifications of the insight that certain actions lead to bad consequences in ways that people consistently fail to predict or appreciate.

All throughout history, various despots and princes have thought "You know, the last hundred times someone tried to restrict freedom of religion, it went badly. Luckily, *my* religion happens to be the One True Religion, and I'm totally sure of this, and everyone else will eventually realize this and fall in line, so *my* plan to restrict freedom of religion will work great!"

Every revolution starts with an optimist who says "All previous attempts to kill a bunch of people and seize control of the state have failed to produce a utopia, but luckily *my* plan is much better and we're totally going to get to utopia this time." Or, as Huxley put it: "Only one more indispensable massacre of Capitalists or Communists or Fascists and there we are - there we are - in the Golden Future."

So another way to put it is that rights don't just say "Doing X has been observed to have bad consequences", but also "Doing X has been observed to have bad consequences, even when smart people are quite certain it will have good consequences."

13.3: Then even though you got to rights by a different route than the libertarians, it sounds like you agree with them that they're inalienable.

It's not as simple as *that* either. Every so often, the conventional wisdom is wrong. So many lunatics and crackpots spent their lives trying to turn lead into gold that it became a classic metaphor for a foolish wild goose chase. The rule "stop trying to transmute elements into each other, it never works" was no doubt a good and wise rule. If more would-be alchemists had trusted this conventional wisdom, and fewer had thought "No, even though everyone else has failed, *I* will be the one to discover transmutation", it would have prevented a lot of wasted lives.

…and then we discovered nuclear physics, which is *all about* transmuting elements into one another, and which works very well and is a vital source of power. And yes, nuclear physicists at Berkeley successfully used a giant particle accelerator to turn lead into gold, although it only works a few atoms at a time and isn't commercially viable.

The point is, the heuristic that you shouldn't waste your life studying transmutation was a good one and very well-justified at the time, but if we had elevated it into a timeless and unbreakable principle, we never would have been able to abandon it after we learned more about nuclear physics and trying to transmute things was no longer so foolish.

Rights are a warning sign that we should not naively expect breaking them to have good consequences. In order to claim even the possibility of good consequences from violating a right, we need to be at least as far away from the actions they were meant to prevent as nuclear physics is to alchemy.

13.3.1: Can you give an example of a chain of reasoning where some government violation of a right is so radically different from the situation that led the right to exist in the first place?

Let's take for example the right that probably dominates discussions between libertarians and non-libertarians: the right to property. On the individual scale, taking someone else's property makes them very unhappy, as you know if you've ever had your bike stolen. On the larger scale, abandoning belief in private property has disastrous results for an entire society, as the experiences of China and the Soviet Union proved so conclusively. So it's safe to say there's a right to private property.

Is it ever acceptable to violate that right? In the classic novel *Les Miserables*, Jean Valjean's family is trapped in bitter poverty in 19th century France, and his nephew is slowly starving to death. Jean steals a loaf of bread from a rich man who has more than enough, in order to save his nephew's life. This is a classic moral dilemma: is theft acceptable in this instance?

We can argue both sides. A proponent might say that the good consequences to Jean and his family were very great - his nephew's life was saved - and the bad consequences to the rich man were comparatively small - he probably has so much food that he didn't even miss it, and if he did he could just send his servant to the bakery to get another one. So on net the theft led to good consequences.

The other side would be that once we let people decide whether or not to steal things, we are on a slippery slope. What if we move from 19th century France to 21st century America, and I'm not exactly starving to death but I really want a PlayStation? And my rich neighbor owns like five PlayStations and there's no reason he couldn't just go to the store and buy another. Is it morally acceptable for me to steal one of his PlayStations? The same argument that applied in Jean Valjean's case above seems to suggest that it is - but it's easy to see how we go from there to everyone stealing everyone's stuff, private property becoming impossible, and civilization collapsing. That doesn't sound like a very good consequence at all.

If everyone violates moral heuristics whenever they personally think it's a good idea, civilization collapses. If no one ever violates moral heuristics, Jean Valjean's nephew starves to death for the sake of a piece of bread the rich man never would have missed.

We need to bind society by moral heuristics, but also have some procedure in place so that we can suspend them in cases where we're exceptionally sure of ourselves without civilization instantly collapsing. Ideally, this procedure should include lots of checks and balances, to make sure no one person can act on her own accord. It should reflect the opinions of the majority of people in society, either directly or indirectly. It should have access to the best minds available, who can predict whether violating a heuristic will be worth the risk in this particular case.

Thus far, the human race's best solution to this problem has been governments. Governments provide a method to systematically violate heuristics in a particular area where it is necessary to do so without leading to the complete collapse of civilization.

If there was no government, I, in Jean Valjean's situation, absolutely would steal that loaf of bread to save my nephew's life. Since there is a government, the government can set a certain constant amount of theft per year, distribute the theft fairly among people whom it knows can bear the burden, and then feed starving children and do other nice things. The ethical question of "is it ethical for me to steal/kill/stab in this instance?" goes away, and society can be peaceful and stable.

13.3.2: So you're saying that you think in this case violating the right will have good consequences. But you just agreed that *even when people think this*, violating the right usually has bad consequences.

Yes, I admit it's complicated. But we have to have some procedures for violating moral heuristics, or else we can't tax to support a police force, we can't fight wars, we can't lie to a murderer who asks us where our friend is so he can go kill her when he finds her, and so on.

The standard I find most reasonable is when it's universalizable and it avoids the issue that caused us to develop the heuristic in the first place.

By universalizable, I mean that it's more complicated than me just deciding "Okay, I'm going to steal from this guy now". There has to be an agreed-upon procedure where everyone gets input, and we need to have verified empirically that this procedure usually leads to good results.

And is has to avoid the issue that caused us to develop the heuristic. In the case of stealing, this is that theft makes property impossible or at least impractical, no one bothers doing work because it will all be stolen from them anyway, and so civilization collapses.

In the case of theft, taxation requires authorization by a process that most of us endorse (the government set up by the Constitution) and into which we all get some input via representative democracy. It doesn't cause civilization to collapse because it only takes a small and extremely predictable amount from each person. And it's been empirically verified to work: as I argued above, countries with higher tax rates like Scandinavia actually *are* nicer places to live than countries with lower tax rates like the United States. So we've successfully side-stepped the insight that stealing usually has bad consequences, even though we recognize that the insight remains true.

13.4: Governments will inevitably make mistakes when deciding when to violate moral heuristics. Those mistakes will cost money and even lives.

And the policy of never, ever doing anything will never be a mistake?

It's very easy for governments to make devastating mistakes. For example, many people believe the US government's War in Iraq did little more than devastate the country, kill hundreds of thousands of Iraqis, and replace Saddam with a weak government unable to stand up to extremist ayatollahs.

But the other solution – never intervening in a foreign country at all – didn't work so well either. Just look at Holocaust-era Germany, or 1990s Rwanda.

Why, exactly, should moral questions be simple?

There is a certain tradition that the moral course of action is something anyone, from the high priest unto the youngest child, can find simply by looking deep in his heart. Anyone who does not find it in his heart is welcome to check the nearest Giant Stone Tablet, upon which are written infallible rules that can guide him through any situation. Intelligence has nothing to do with it. It should be blindingly obvious, and anyone who claims it has a smidgen of difficulty or vagueness is probably an agent of the Dark Lord, trying to seduce you from the True Path with his lies.

And so it is tempting to want to have some really easy principle like "Never get involved in a foreign war" and say it can never lead you wrong. It makes you feel all good and warm and fuzzy and moral and not at all like those evil people who don't have strong principles. But real life isn't that simple. If you get involved in the wrong foreign war, millions of people die. And if you don't get involved in the right foreign war, millions of people also die.

So you need to have good judgment if you want to save lives and do the right thing. You can't get a perfect score in morality simply by abdicating all responsibility. Part of the difficult questions that all of us non-libertarians have been working on is how to get a government that's good at answering those sorts of questions correctly.

13.5: No, there's a difference. When you enter a foreign war, you're killing lots of people. When you don't enter a foreign war, people may die, but it's not your job to save them. The government's job is only to protect people and property from force, not to protect people from the general unfairness of life.

Who died and made you the guy who decides what the government's job is? Or, less facetiously: on what rational grounds are you making

that decision?

Currently, several trillion dollars are being spent to prevent terrorism. This seems to fall within the area of what libertarians would consider a legitimate duty of government, since terrorists are people who initiate force and threaten our safety and the government needs to stop this. However, terrorists only kill an average of a few dozen Americans per year.

Much less money is being spent on preventing cardiovascular disease, even though cardiovascular disease kills 800,000 Americans per year.

Let us say, as seems plausible, that the government can choose to spend its money either on fighting terrorists, or on fighting CVD. And let us say that by spending its money on fighting terrorists, it saves 40 lives, and by spending the same amount of money on fighting CVD, it saves 40,000 lives.

All of these lives, presumably, are equally valuable. So there is literally no benefit to spending the money on fighting terrorism rather than CVD. All you are doing is throwing away 39,960 lives on an obscure matter of principle. It's not even a good principle – it's the principle of wanting to always use heuristics even when they clearly don't apply because it sounds more elegant.

There's a reason this is so tempting. It's called the Bad Guy Bias, and it's an evolutionarily programmed flaw in human thinking. People care much more about the same amount of pain when it's inflicted by humans than when it's inflicted by nature. Psychologists can and have replicated this in the lab, along with a bunch of other little irrationalities in human cognition. It's not anything to be ashamed of; everyone's got it. But it's not something to celebrate and raise to the level of a philosophical principle either.

13.6: Stop calling principles like "don't initiate force" heuristics! These aren't some kind of good idea that works in a few cases. These are the very principles of government and morality , and it's literally impossible for them to guide you wrong!

Let me give you a sketch of one possible way that a libertarian perfect world that followed all of the appropriate rules to the letter could end up as a horrible dystopia. There are others, but this one seems most black-and-white.

Imagine a terrible pandemic, the Amazon Death Flu, strikes the world. The Death Flu is 100% fatal. Luckily, one guy, Bob, comes up with a medicine that suppresses (but does not outright cure) the Death Flu. It's a bit difficult to get the manufacturing process right, but cheap enough once you know how to do it. Anyone who takes the medicine at least once a month will be fine. Go more than a month without the medicine, and you die.

In a previous version of this FAQ, Bob patented the medicine, and then I got a constant stream of emails saying (some) libertarians don't believe in patents. Okay. Let's say that Bob doesn't patent the medicine, but it's complicated to reverse engineer, and it would definitely take more than a month. This will become important later.

Right now Bob is the sole producer of this medicine, and everyone in the world needs to have a dose within a month or they'll die. Bob knows he can charge whatever he wants for the medicine, so he goes all out. He makes anyone who wants the cure pay one hundred percent of their current net worth, plus agree to serve him and do anything he says. He also makes them sign a contract promising that while they are receiving the medicine, they will not attempt to discover their own cure for the Death Flu, or go into business against him. Because this is a libertarian perfect world, everyone keeps their contracts.

A few people don't want to sign their lives away to slavery, and refuse to sign the contract. These people receive no medicine, and die. Some people try to invent a competing medicine. Bob, who by now has made a huge amount of money, makes life extremely difficult for them and bribes biologists not to work with them. They are unable to make a competing medicine within a month, and die. The rest of the world promises to do whatever Bob says. They end

up working as peons for a new ruling class dominated by Bob and his friends.

If anyone speaks a word against Bob, they are told that Bob's company no longer wants to do business with them, and denied the medicine. People are encouraged to inform on their friends and families, with the promise of otherwise unavailable luxury goods as a reward. To further cement his power, Bob restricts education to the children of his friends and strongest supporters, and bans the media, which he now controls, from reporting on any stories that cast him in a negative light.

When Bob dies, he hands over control of the medicine factory to his son, who continues his policies. The world is plunged into a Dark Age where no one except Bob and a few of his friends have any rights, material goods, or freedom. Depending on how sadistic Bob's and his descendants are, you may make this world arbitrarily hellish while still keeping perfect adherence to libertarian principles.

Compare this to a similar world that followed a less libertarian model. Once again, the Amazon Death Flu strikes. Once again, Bob invents a cure. The government thanks him, pays him a princely sum as compensation for putting his cure into the public domain, opens up a medicine factory, and distributes free medicine to everyone. Bob has become rich, the Amazon Death Flu has been conquered, and everyone is free and happy.

13.6.1: This is a ridiculously unlikely story with no relevance to the real world.

I admit this particular situation is more a *reductio ad absurdum* than something I expect to actually occur the moment people start taking libertarianism seriously, but I disagree that it isn't relevant.

The arguments that libertarianism will protect our values and not collapse into an oppressive plutocracy require certain assumptions: there are lots of competing companies, zero transaction costs, zero start-up costs, everyone has complete information, everyone has free choice whether or not to buy any particular good, everyone behaves

rationally, et cetera. The Amazon Death Flu starts by assuming the opposite of all of these assumptions: there is only one company, there are prohibitive start-up costs, a particular good absolutely has to be bought, et cetera.

The Amazon Death Flu world, with its assumptions, is not the world we live in. But neither is the libertarian world. Reality lies somewhere between the "capitalism is perfect" of the one, and the "capitalism leads to hellish misery" of the other.

There's no Amazon Death Flu, but there are things like hunger, thirst, unemployment, normal diseases, and homelessness. In order to escape these problems, we need things provided by other people or corporations. This is fine and as it should be, and as long as there's a healthy free market with lots of alternatives, in most cases these other people or corporations will serve our needs and society's needs while getting rich themselves, just like libertarians hope.

But this is a contingent fact about the world, and one that can sometimes be wrong. We can't just assume that the heuristic "never initiate force" will *always* turn out well.

13.7: The government doesn't need to violate moral heuristics. In the absence of government programs, private charity would make up the difference.

Find some poor people in a country without government-funded welfare, and ask how that's working out for them.

Private charity from the First World hasn't prevented the Rwandans, Ethiopians, or Haitians from dying of malnutrition or easily preventable disease.

It's possible that this is just because we First Worlders place more importance on our own countrymen than on foreigners, and if Americans were dying of malnutrition or easily preventable disease, patriotism would make us help them.

The US government currently spends about $800 billion on welfare-type programs for US citizens. Americans give a total of $300 billion to charity per year.

Let's assume that private charity is twice as efficient as the government (in reality, it's probably much less, since the government has economies of scale, but libertarians like assumptions like this and I might as well indulge them).

Let's also assume that only half of charity goes to meaningful efforts to help poor American citizens. The other half would be things like churches, the arts, and foreign countries.

Nowadays, a total of $550 billion (adjusted, govt+private) goes to real charity (800b*1/2+300b*1/2). If the government were to stop all welfare programs, this number would fall to $150 billion (adjusted). Private citizens would need to make up the shortfall of $400 billion to keep charity at its current (woefully low) level. Let's assume that people, realizing this, start donating a greater proportion (66%) of their charity to the American poor instead of to other causes. That means people need to increase their charity to about $830 billion ([400b + 150b]/.66).

Right now, 25% is a normal middle-class tax rate. Let's assume the government stopped all welfare programs and limited itself to defense, policing, and overhead. There are a lot of different opinions about what is and isn't in the federal budget, but my research suggests that would cut it by about half, to lower tax rates to 12.5%.

So, we're in the unhappy situation of needing people to almost triple the amount they give to charity even though they have only 12.5% more money. The real situation is much worse than this, because if the government stopped all programs except military and police, people would need to pay for education, road maintenance, and so on out of their own pocket.

My calculations are full of assumptions, of course. But the important thing is, I've never seen libertarians even try to do calculations. They just assume that private citizens would make up the shortfall. This is the difference between millions of people leading decent lives or starving to death, and people just figure it will work out without checking, because the free market is always a Good Thing.

That's not reason, even if you read it on www.reason.com. That's faith.

13.8: People stupid enough to make bad decisions deserve the consequences of their actions. If government bans them from making stupid decisions, it's just preventing them from getting what they deserve.

One of my favorite essays, Policy Debates Should Not Appear One-Sided, provides a much better critique of this argument than I could. It starts by discussing a hypothetical in which the government stopped regulating the safety of medicines. Some quack markets sulfuric acid as medicine, and a "poor, honest, not overwhelmingly educated mother of five children" falls for it, drinks it, and dies.

If you were really in that situation, would you really laugh, say "Haha, serves her right" and go back to what you were doing? Or would it be a tragedy even though she "got what she deserved"?

The article ends by saying:

> *Saying 'People who buy dangerous products deserve to get hurt!' is not tough-minded. It is a way of refusing to live in an unfair universe. Real tough-mindedness is saying, 'Yes, sulfuric acid is a horrible painful death, and no, that mother of 5 children didn't deserve it, but we're going to keep the shops open anyway because we did this cost-benefit calculation.'...I don't think that when someone makes a stupid choice and dies, this is a cause for celebration. I count it as a tragedy. It is not always helping people, to save them from the consequences of their own actions; but I draw a moral line at capital punishment. If you're dead, you can't learn from your mistakes.*

Read also about the just-world fallacy. "Making a virtue out of necessity" shouldn't go as far as celebrating deaths if it makes your political beliefs more tenable.

Part E: Practical Issues

The Argument: *Allowing any power to government is a slippery slope toward tyranny. No matter what the costs or benefits of any particular proposal, libertarians should oppose all government intrusion as a matter of principle.*

The Counterargument: *This fundamentally misunderstands the ways that nations collapse into tyranny. It also ignores political reality, and it doesn't work. Libertarians should cooperate with people from across the ideological spectrum to oppose regulations that doesn't work and keep an open mind to regulation that might.*

14. Slippery Slopes

14.1: I'm on board with doing things that have the best consequences. And I'm on board with the idea that *some* government interventions *may* have good consequences. But allowing any power to government is a slippery slope. It will inevitably lead to tyranny, in which do-gooder government officials take away all of our most sacred rights in order to "protect us" from ourselves.

History has *never* shown a country sinking into dictatorship in the way libertarians assume is the "natural progression" of a big-government society. No one seriously expects Sweden, the United Kingdom, France, or Canada to become a totalitarian state, even though all four have gone much further down the big-government road than America ever will.

Those countries that have collapsed into tyranny have done so by having so *weak* a social safety net and so *uncaring* a government that the masses felt they had nothing to lose in instituting Communism or some similar ideology. Even Hitler gained his early successes by pretending to be a champion of the populace against the ineffective Weimar regime.

Czar Nicholas was not known for his support of free universal health care for the Russian peasantry, nor was it Chiang Kai-Shek's attempts to raise minimum wage that inspired Mao Zedong. It has generally been among *weak* governments and a *lack* of protection

for the poor where dictators have found the soil most fertile for tyranny.

14.1.1: But still, if we let down our guard, bureaucrats and politicians will have free rein to try to institute such a collapse into dictatorship.

I have always found the libertarian conviction that all politicians are secretly trying to build up their own power base to *1984*-ish levels a bit weird.

All the time, I am hearing things like "No one really believes in global warming. It's just a plot by the government to expand control over more areas of your life." Or "since private charity is a threat to government's domination of social welfare, once government gets powerful enough it will try to ban all private charity."

Sure, people really do like power. But usually it's the sort of power that comes with riches, fame, and beautiful women willing to attend to your every need. Just sitting in your office, knowing in an abstract way that because of you a lot of people who might otherwise be doing useful industry are fretting about their carbon emissions - that's not the kind of power people sell their souls for. The path to ultimate domination of all humanity does not lead through the Dietary Fiber Levels in Food Act of 2006.

Most folk like to think of themselves as good people. Sure, they may take a bribe or two here, and have an affair or two there, and lie about this and that, "but only for the right reasons." The thought process "Let me try to expand this unnecessary program so I can bathe in the feeling of screwing American taxpayers out of more of their hard-earned money" is not the kind that comes naturally, especially in a society where it leads to minimal personal gain. A politician who raises your taxes can't use the money to buy himself a new Ferrari. At least, he can't do it directly, and if he really wants that Ferrari there have got to be much easier ways to get it.

Human beings find it hard to get angry at a complicated system, and prefer to process things in terms of evil people doing evil things.

Eliezer Yudkowsky of Less Wrong writes:

> *Suppose that someone says "Mexican-Americans are plotting to remove all the oxygen in Earth's atmosphere." You'd probably ask, "Why would they do that? Don't Mexican-Americans have to breathe too? Do Mexican-Americans even function as a unified conspiracy?" If you don't ask these obvious next questions when someone says, "Corporations are plotting to remove Earth's oxygen," then "Corporations!" functions for you as a semantic stopsign.*

And if you don't ask some of these same questions when someone says "Government wants to take away freedom!," then you're not thinking of government as a normal human institution that acts in normal human ways.

15. Strategic Activism

15.1: All you've argued so far is that it's possible, in theory, for an *ideal* government making some very clever regulations to do a little more good than harm. But that doesn't prove that the *real* government does more good than harm, and in fact it's probably the opposite. So shouldn't we admit that in a hypothetical perfect world government might do some good, while still being libertarians in reality?

I think if you've got enough intelligence and energy to be a libertarian, a better use of that intelligence and energy would be to help enact a properly working system.

15.2: It's impossible to improve government; because power corrupts, all conceivable forms of government will be ineffective, wasteful, and dishonest.

"Impossible" is a really strong word.

Economist Robin Hanson has a proposal for a market-based open-source form of government called "futarchy", in which government policies are decided entirely by a prediction market. Prediction markets operate similarly to stock markets and allow participants to

buy or sell shares in predictions - for example, a share that pays out $100 if the economy improves this year, but $0 if the economy deteriorates. If it settles around a price of $60, this means the investing public predicts as 60% chance that the economy will go up.

A prediction market could be used to set policy by predicting its effects: for example, by comparing the prices of "we will institute the president's economic plan, and the economy will improve" , "we will not institute the president's economic plan, and the economy will improve" and "we will institute the president's economic plan", we can determine the public's confidence that the president's plan will improve the economy. There are some nifty theorems of economics that prove that such a market would produce a more accurate estimate of the plan's chances than any other conceivable method (including consulting experts), and that it would be very difficult to corrupt. You can read more about it here.

My point isn't that futarchy would definitely work. It's that it's an example of some of the best ideas that smart people trying to improve government can come up with. And unless you're creative enough to develop futarchy on your own, or well-read enough to be sure you've heard of it and everything else like it, you're being premature in calling improvements in government "impossible".

15.3: Even if there are ways to improve government, they are impractical because they're too politically unpopular.

Let's be totally honest here. The US Libertarian Party currently has a grand total of zero state legislators, zero state governors, zero representatives, and zero senators. It's never gotten much above one percent in any presidential election. Nor have any successful or nationally known major-party candidates endorsed genuinely libertarian ideals except maybe Ron Paul, who just suffered his third landslide defeat.

The libertarian vision of minimal government is politically impossible to enact. This is not itself an argument against it - most

good ideas are - but it does mean you can't condemn the alternatives for being politically impossible to enact.

Incremental attempts to improve government have a much better track record, both in terms of political palatability and success rate, than libertarian efforts to dismantle government whole-cloth. If you want to focus on something that might work, you should concentrate your efforts there.

15.4: Isn't it better to draw a line in the sand and say no government intervention at all? This keeps us off the slippery slope to the kind of awful, huge government we have today.

Empirically, no. Again I point out that libertarianism has been completely ineffective as a political movement. The line-in-the-sand idea is an interesting one but obviously hasn't worked.

And there are some serious advantages to erasing it. If non-libertarians see libertarians as ideologues who hate all government programs including the ones that could work, then they will dismiss any particular libertarian objection as meaningless: why pay attention to the fact that a libertarian hates this particular bill, when she hates *every* bill?

But if libertarians took a principled stand in favor of some government regulation that might work, they could credibly say "Look, it's not that we have a knee-jerk hatred for all possible regulations, it's just that *this particular regulation is a horrible idea*." And people might listen.

It *might* also help arrest the polarization of society into factions who apply ideological "litmus tests" to all proposals before even hearing them out (eg pretty much all self-described "progressives" will automatically support any proposal to be tougher on pollution without even looking at what the economic costs versus health benefits will be, and most self-described libertarians will automatically oppose it just as quickly.) This sort of thing needs to stop, libertarians are one of the at least two groups who need to stop

it, and the more people who stop, the more people on both sides will notice what they're doing and think about it a little harder.

16. Miscellaneous and Meta

16.1: I still disagree with you. How should I best debate you and other non-libertarians in a way that is most likely to change your mind?

The most important advice I could give you is don't come on too strong. Words like "thievery" and "enslave" are emotional button pressers, not rational arguments. Attempts to insult your opponents by calling them tyrants or suggesting they want to rule over the rest of humanity as slaves and cattle (yes, I've gotten that) is more likely to annoy than convince. And please, stop the "1984" references, especially when you're talking about a modern liberal democracy. Seriously. It's like those fundamentalists who have websites about how not having prayer in school is equivalent to the Holocaust.

Many non-libertarians aren't going to be operating from within the same moral system you are. Sometimes the libertarians I debate don't realize this and this causes confusion when they try to argue that something's morally wrong. If you want to convince your opponent on moral grounds, you're either going to have to show how their theories fail *even by their own moral standards*, or else prove your standards are right by deriving them from first principles (warning: this might be impossible).

Don't immediately assume that just because we are not libertarians, we must worship Stalin, love communism, think government should be allowed to control every facet of people's lives, or even support things like gun control or the War on Drugs. Non-libertarianism is a lot like non-Hinduism: it's a pretty diverse collection of viewpoints with everything from full-on fascists to people who are totally libertarian except about one tiny thing.

Finally, you may have better luck convincing us of specific points, like "Government should not set a minimum wage" than broad

slogans, like “Government can never do anything right.” It’s *really* hard to prove a universal negative.

16.2: Where can I go to see a rebuttal to this FAQ?

So far there is only one rebuttal I know about, which is based on a previous version and therefore sort of obsolete. You can find it here: Why You Shouldn’t Hate My Freedom.

If you’ve written another rebuttal or you know of one, email me (address below) and I’ll add it here.

16.3: Where can I go to find more non-libertarian information?

Mike Huben has a *terrifyingly* large collection of non-libertarian and anti-libertarian material of wildly varying quality and tone at his website.

16.4: How can I respond to this FAQ in some way?

My email is scott period siskind at-symbol gmail period com. Feel free to email me if you have any questions, complaints, or comments on this FAQ. Although do I try to read all the email I receive, after getting more than I expected from previous versions of this document I am going to concede defeat and admit I probably won’t respond to every letter.

A Blessing in Disguise, Albeit a Very Good Disguise

Very many of my friends sing the praises of modafinil (I have not tried it myself). They say it can make you more focused, more productive, and at least temporarily remove the basic human need for sleep. It doesn't have the normal stimulant side effects of "buzz" and agitation. And it's cheap and has fewer side effects than aspirin (EDIT: it does interact with other drugs including birth control and should be used with caution if you're on anything else; thank you celandine13).

(it's really convenient that aspirin became a poster child for "safe, commonly used medication" despite having such an crazy array of potential deadly side effects. It means that whenever you want to push a new drug, you can say it has "fewer side effects than aspirin" and be pretty sure that you're right)

Despite its excellent safety profile, it is currently a Schedule IV controlled substance in the United States.

Doesn't this mean that I must be wrong about its excellent safety profile? No. See for example Gwern's research on the subject. About half the people reading this paragraph are going to say "Wait, don't the FDA and the entire decision-making apparatus of the United States government have more data and credibility than one guy with a website?" The other half of the people know Gwern.

It's also worth noting that adrafinil, a prodrug of modafinil which is strictly more dangerous because it contains all the

side effects of the latter plus a risk of liver damage, is totally legal without any prescription at all. And modafinil is freely available over the counter in various countries (I think Spain and India) and they have yet to collapse into unspeakable wastelands of despair.

(actually, Spain kinda did, but it seems unrelated.)

It's *also* worth noting that the alternative to modafinil is using legal stimulants, like Red Bull and Four Loco. These actually *are* dangerous and can, for example, cause abnormal heart rhythms that kill you. We also saw a steady trickle of energy drink overusers in the psychiatric hospital, and although you probably need to have an inborn disposition for energy drinks to tip you over the edge, who knows how common such an inborn disposition really is? Modafinil is probably way safer than these totally unrestricted alternatives.

So you would think that I am going to argue that modafinil should be legalized. Or at least that the cultural stigma against using it should be relaxed. But that would be too easy. Actually, I want to argue the opposite.

Let's assume that the wildest claims of my friends are correct. Some of these friends got through medical school with relatively little damage by using modafinil to study eight or ten hours a day and skip sleep. Others are in the rationality community and use it to concentrate on their programming or mathematics work. They mostly agree with Gwern that it can be modeled as adding four hours to the day, both in the form of costlessly lost sleep and in the form of greater attention during waking hours.

Economists distinguish between positional goods and…and I can't find what the opposite of a positional good is, so let's call it an absolute good. A positional good is something where it doesn't matter exactly how much of the good you have, but only what your ranking is relative to everyone else. Superyachts are probably a positional good. I don't think anyone thinks "Man, this 100 foot yacht is crap, there isn't nearly enough room for all my yachting-related activities." They think "My neighbor has a 200 foot yacht; my 100 foot yacht looks crappy in comparison. I should build a 300 foot yacht." If the person involved had the option of destroying her neighbor's 200 foot yachts, then her 100 foot yacht would suddenly become more than enough.

An absolute good is the opposite. For example, if you're injured, you want painkillers as an absolute good. It doesn't matter whether your neighbors are getting more or less painkillers than you are, so much as that you are getting enough painkillers to take away your pain.

Except it's actually really hard to think of pure absolute goods. A lot of things I was going to put as my absolute good example don't really work, because our idea of what's acceptable is set by our friends and neighbors. In Haiti, people who had a house made of real sheet metal felt awesome, because most of their neighbors were still living in refugee tents; meanwhile in America a house made of sheet metal would be awful because everyone needs to have a McMansion; a McMansion, however, is quite sufficient. But in the postsingularity thoughtspace of 19-uvara-46-asxura, everyone has their own continent perfectly terraformed as a projection of ver innermost dreams, and someone with a McMansion feels as left-out and squalid as someone living in

a sheet-metal shack in America.

The richer you are, the more your goods shift from absolute to positional; 90% of the value of a $5000 used car is its getting-you-places-ness, but 90% of the value of a $500000 Ferrari is its looking-cool-relative-to-other-cars.

Right now America and to a lesser degree other first-world countries are caught in a trap where almost all of their economic growth is funneled to the rich and upper-middle-class, who spend it on positional goods. Since all the rich people are spending it on positional goods equally, none of their relative position changes in any interesting way and all of the positional goods are useless.

Therefore in the modern era most economic growth in first-world countries is pretty useless as a direct action. There may be useful indirect actions, like advancing technology, increasing tax revenue that can be spent on useful absolute goods, and increasing the amount that flows as charity to the Third World, but the actual direct effect of economic growth is pretty close to zero.

Okay, let's go back to modafinil. Right now the FDA is pretty incompetent and doesn't enforce any of its own restrictions, so in practice anyone can get modafinil. And getting modafinil is currently very useful. If you're in medical school, and you're not doing very well, you can take some modafinil, gain a big unfair advantage over your peers, and shoot up the class rankings. If you're an executive, you can work much harder and get a promotion your friends can't. If you're a programmer, you can amaze the world with your vastly improve programming output.

But let's say the FDA restrictions on modafinil switched from "poorly enforced" to "nonexistent", and let's say that at the same time the cultural stigma against using mind-enhancing drugs went away. Now what?

Now instead of hiding their use behind vague rumors, those medical students trumpet their brilliant discovery of this new wonder drug to everyone. All medical students start taking modafinil, except maybe some with religious restrictions or something. Of course, this doesn't mean that all medical students get As all the time. It means that the medical schools make their coursework much harder, and the medical students go back to being on the cusp of failure. Except now that it's harder, it's impossible for most students to pass medical school without modafinil. So the religious people flunk out, everyone else has to work much harder, and in the end no student gains. *Arguably* future patients might gain from having better trained doctors, but I think this wildly overestimates the usefulness of the medical education system.

The same is true of executives. Now modafinil no longer means an easy promotion. Now all the executives start taking modafinil, and everyone has the same chance of getting promoted as before, except the religious people and the people who are allergic to modafinil and anyone who has a personal preference for getting more than three hours of sleep per night even though it's not strictly necessary.

Basically, obligations are a demon that eats up all the free time and happy things in your life. If only a few people have modafinil, they have an extra weapon against the demon. If everyone has modafinil, expectations and competition increase

and so the demon becomes stronger. A new equilibrium is established in which there's more economic growth (so the rich get some more useless positional goods) but everyone gets four hours less sleep per night, plus they have to spend money on modafinil, plus the few people who can't take modafinil for one reason or another are screwed.

"But wait," you say. "Couldn't people just decide to work shorter hours and instead use the extra time they have in the day to see their family or pursue their hobbies or volunteer or do something *good*?"

Yes, we don't live in a totalitarian society, so that choice technically exists. Just as the choice *technically* exists for people to try that now. Most people earn much more than they need to live. So *in theory*, they have the option of working twenty-five hour weeks and spending the extra fifteen hours hiking or gardening.

But in practice, people don't. The majority of well-paying acceptable jobs demand a forty-hour work week, and most people don't have the freedom to look for the ones that don't. It costs companies less money in training and overhead to hire two people to work 40-hour weeks than four people to work 20-hour weeks, and so they will always prefer the 40-hour workers. If you want to be a prestigious doctor or lawyer or executive or whatever you have to signal your commitment by working even *longer* than the 40-hour Schelling point. In practice, you're working as long as the companies are legally and socially permitted to make you, which in our society is 40 hours.

If suddenly days magically get four more hours in them, then

the work week will shoot up to 60 hours and stay there. People might get paid more, but the economy will adjust so that the extra money becomes necessary just to tread water, the same way it looked like people were getting paid more when women entered the work force and the family could theoretically double its income but everything adjusted. The extra economic growth will go to positional goods for the rich, and you will get 20 hours less sleep per week (granted without a corresponding decrease in restfulness), have to pay for modafinil out of your own pocket, but otherwise be in about the same position.

(couldn't the government just make a law fixing the work week at its current length thus preventing this race to the bottom and all of its unfortunate consequences? In an ideal world, yes, but the small-L libertarians would never allow it.)

So legalizing modafinil (with corresponding reduction of stigma) leads directly to you having to work four hours more every day, gain an extra item on your budget (modafinil: $1000-$3000/year), get four hours less sleep (admittedly without restfulness cost, but still unpleasant especially for a lucid dreaming hobbyist like myself), plus suffer any unknown side effects of the drug that might turn up. And for all this, you get the chance to earn money that the economy immediately siphons off and throws away on more positional goods.

Despite this I'm still not sure it would be so bad. Economic growth is a pretty powerful force, and even if most of the force is wasted there are still those small direct effects on the poor/middle class plus the indirect effects which might end up being much more powerful. And maybe the government will stand up to the libertarians and fix the work-week, or the

creeping increase wouldn't be as inevitable as I think.

But compare these possible benefits of legalization to how downright *optimal* the current modafinil regime is.

From what two of my friends in the modafinil business have told me, it's really easy to get modafinil now - just order it online with PayPal and wait a little while for shipping. And no one ever really gets in trouble for it; Gwern's research turns up only a single case in the entire history of the US in which someone got busted for modafinil, and he speculates it was just a racist Southern court looking for some excuse to convict a poor suspicious-looking black person. This probably does not generalize to risk for the average user.

So in practice, the current regime offers no downsides to seeking modafinil. It is much more of a psychological barrier than an actual barrier. But it is an effective psychological barrier, which only a few people get across. Who?

First of all, they have to be individuals rather than institutions. A big Fortune 500 company requiring all of its employees to take modafinil probably *would* get busted by the FDA.

Second, they can't care too much about social stigma. There's still a stigma on stimulant use, probably carried over from some of the other stimulants which really are pretty scary (WAItW, anyone?). And of course there's a stigma on breaking the law.

Third, they have to be intelligent. Anyone without at least a little curiosity is going to do what everyone else is doing and take Red Bull or Four Loco. They're never going to find good

analyses like Gwern's research, and they probably couldn't understand them even if they did. An unintelligent person won't be able to distinguish modafinil from the thousand different quack remedies that are supposed to make you "more awake!" and "give you the extra energy you need to complete your day!"

Fourth, they have to be kind of..not really anti-establishment, but at least less violently pro-establishment than usual. It's pretty hard for most people to say "Well, I guess the government is wrong about this, might as well circumvent them." But that's pretty much all the counter-culture ever *does*.

So: individual intelligent non-social anti-establishment people. Basically geeks. And a very specific kind of geek, too. I won't specify exactly which kind beyond that link, because internecine geek feuds always turn ugly, but I think it is pointing to a particular geek cluster.

It's hard not to be suspicious that God has planned this all along. He's basically saying "Behold, geeks, you are My chosen people, so I give unto you a major advantage over non-geeks. The hilarious part shall be that it is self-selecting; anyone who chooses to use this is the sort of person I trust to have an advantage in society. Anyone who chooses not to use it, well, they probably would just screw it up anyway."

And because these geeks remain a very small percent of the population, the problems with large-scale use don't occur. The angel Technology giveth with the right hand, but the demon Economics doesn't notice and so doesn't wake up to taketh away with the left hand. It's not that it's a win-win situation. It's that it's a win-neutral situation, which in terms of

positional goods is even better.

So what does it say about me that I don't (haven't yet?) used modafinil? I'm not sure. I've always known I'm not a very good anti-establishment specific-cluster geek. Last night when a friend was explaining his theory of PCs (people who are actively doing interesting work and changing the world in such a way that things revolve around them) and NPCs (people who mostly just hang around and provide background), I might have been the only person at the table not especially convinced he was a PC.

Not that I feel any deep sense of inadequacy about this. NPCs can be pretty neat too. Schala was an NPC. If I can be as awesome as she was, I'll be pretty happy.

Oh, right. Nothing in this post should be taken as any kind of official medical endorsement of modafinil, which I have not studied in a medical context and which I am not anywhere near officially qualified to recommend or disrecommend. Nothing else in this post was more than about 60% serious, but this paragraph is *entirely* serious.

[**EDIT**: 60% serious may have been an overestimate (or we may have different scales of seriousness percent). I think the argument is correct in saying the benefits from modafinil would be much lower than most people think, but I was not entirely serious in saying they would be zero, or less than the costs. I would, with some trepidation and a high expectation of regretting it later, endorse legalizing modafinil]

Basic Income Guarantees

Basic income guarantees.

The first time I heard about them was five years ago, and I decided they were stupid. I think I thought about them again briefly two or three years ago, and was still pretty sure they were stupid. A couple weeks ago, wallowinmaya from Less Wrong asked me what I thought about them, and I was all prepared to say they were *still* stupid, but after thinking about it longer I'm not so sure.

A basic income guarantee is a system where the government pays everyone in the country a small but liveable income, let's say $15000. If you're poor, you get $15000 a year to live on. If you're rich, you get $15000 from the government above and beyond what you earn from your corporate empire. Everyone in the country, rich or poor, employed or unemployed, young or old, gets $15000.

And the obvious reason it's stupid is that someone has to pay for that. And giving every US adult $15000 a year would cost somewhere around the order of $4 trillion, or just over the current Federal budget.

The real cost would be a bit less, because the government could save some money on things like welfare payments now that nobody is *really* all that poor. But it would be pretty hard to imagine it costing less than $3 trillion or so, meaning we'd have to *at least* double taxes, which would have all sorts of horrible domino effects.

And there is much for everyone to hate about the proposal. If you're the type who doesn't like welfare because it takes money away from productive people and gives it to unproductive people who might not even be trying that hard, well, basic income guarantee does the same thing, only much more so. And if you do like welfare, because you think it's important to help the poor, well, basic income guarantee takes the vast majority of the money it raises and hands it over to the middle-class and rich, making them richer. If you're going to give the government ungodly amounts of money to distribute, why not reserve it for people who really need it?

And although the optimist in me conceives of people who use their newfound freedom from fear of poverty to pursue the careers they've always dreamed of as musicians or inventors, or to live in the forest in harmony with nature, the realist in me knows that the vast majority of those people would in fact spend their time drinking beer and watching TV and having ten kids who they never send to school because *obviously* if you don't need literacy for a job later attaining it just wastes valuable reality-show-watching-time.

So those were the reasons I used to think basic income guarantees were stupid. The reason I'm not so sure now involves structural unemployment and the idea of post-scarcity society.

Back in the 50s, everyone assumed robots would be doing all our work by now and we'd be sitting by the beach all day sipping robot-stirred martinis. That never happened, but it wasn't entirely the roboticists' fault: we *did* automate a lot of formerly difficult jobs. It just turned out that instead of the people whose jobs were replaced by robots sitting on the

beach all day drinking martinis, they become unemployed and essentially unemployable since their only skills were things robots could do better. Although “Well, they should retrain” is a nice thought, not every 50 year old grizzled miner can learn how to program social networking software. So most of them just became destitute and miserable. The gains from automating manufacturing went partly to people in nonmanufacturing fields, who could get more manufactured goods at cheaper prices, and to rich people who owned manufacturing companies and managed to cut costs.

In the future, we can expect technology to replace more and more jobs. This isn’t just in the sense of dominating entire job categories like auto manufacturing (although they’ll do that too - secretaries and waiters won’t be long for this world once we get voice recognition and mobility at low costs) but even in terms of making jobs easier - so that now one engineer can do the work it used to take two engineers, with the second engineer out of a job. The winners will smart people, who can get jobs in technology, and rich people, who can invest in technology and sell what it produces. The losers will be all the unemployed people.

Extending the trend out into the far future and potentially past the singularity, humans will be relatively useless for all forms of work, including robot design (by that time we’ll have robot-designing robots). The only people with access to any wealth will be people who own technology and live off what it produces. This is quite like the feudal economy where if you were born owning land you could live off it forever with no work, and if you were born without land, you were out of luck.

This is a relatively dystopian future - enough technology to

give everyone a fantastic standard of living with minimal work, but the majority of people being poor and miserable because the technology is concentrated in the hands of a few people who have no incentive to share it with anyone else.

(if you think society is too smart to fall for this, it's essentially the situation right now with world hunger. We have more than enough land/technology/etc to feed everyone in the world, but the poor can't afford food and no farmers want to produce food for free, so the technology goes to making silly luxuries for rich people like sunglasses for dogs. The poor can and do break out of their condition through having natural and human resources that the rich want and will trade for, but as technology increases this advantage will disappear.)

As I write this, this sounds sort of Marxist with stuff about the means of production and so on. But Marx was wrong for a few reasons. For one, workers could save up to own the means of production themselves. For another, human capital proved to be more important than machinery during his era. For a third, the capitalists needed the workers almost as much as the workers needed the capitalists, and advances in worker organization and state regulation gave the workers more bargaining power. In a society where labor becomes less valuable, or completely useless, these checks on the Marxist system disappear.

This whole spiel about technology displacing workers isn't just for the far future. Some economists have suggested this is going on now - that the banking crisis certainly didn't help, but that a lot of ther reason unemployment is so high now is that the economy just really doesn't need that many unskilled people any more, and not everyone has (or can develop) skills.

I don’t see an economic or scientific pathway from here to the future where we’re all sitting on the beach enjoying the fruits of technology, as opposed to the future where everyone’s unemployed and poor except the people who own the technology. The only path I can think of is a political one, in which we start redistributing the heck out of income. And simple welfare won’t work; a world in which everyone is on the dole and being constantly hounded by welfare officers and looked down upon by the few people with paying jobs is almost as dystopian as the one where everyone starves to death. At some point we have to say that most people can’t produce wealth and that’s okay.

It may be too early to start such a redistribution program, although depending on how the economic indicators turn out it might not be. But I would feel a whole lot better if society was at least discussing this question and had a good plan for the transition to a post-labor stage.

Book Review: The Nurture Assumption

The latest book I read was *The Nurture Assumption* by Judith Rich Harris, which was supposed to argue that parents don't really have much of an effect on how their kids turn out.

This sounded ridiculous when I first heard it, but people I trusted like Steven Pinker kept endorsing it, so I finally picked it up. The thesis might be a little more subtle than that. Parents can still impact their kids' biological development - to take an extreme example, if you malnourish a baby, that's going to hurt brain development. They can still guide them into certain areas by, again to take an extreme example, making them go to music lessons every day starting at age four. But they don't have to worry that by being too strict or not strict enough or just the right amount of strict but at the wrong time they're going to seriously harm their children's adult personalities. The most dutiful helicopter parents probably wouldn't change much by plopping their kids on the couch every day and telling them not to bother them.

The evidence is pretty overwhelming. The best support comes from studies of identical twins vs. identical twins separated at birth vs. fraternal twins vs fraternal twins separated at birth. These find that about 50% of the variation in personality is genetic (actually, pretty much every study on personality seems to converge around this number) and the other half is not-genetic. But the not-genetic half has nothing to do with parenting - identical twins raised by the same parents have just as many not-gentic differences as identical twins raised apart, and the same is true of fraternal twins. So half of the difference in the way kids turn out is genetic, but the other half

isn’t related to parenting.

Scientists have been slow to accept these findings because they have a bunch of opposing studies that match parenting style to results. But Harris does a beautiful dissection of these studies, a dissection pretty illustrative for anyone who has too much trust in the modern scientific process. For example, studies do show that parents who adhere very meticulously to the standard parenting advice have children who, let’s say, do better at school. But Harris points out - what personality trait is necessary to adhere meticulously to the latest parenting fads? Conscientiousness. What personality trait is necessary to do well at school? Conscientiousness. And what personality trait is about 50% heritable (recall that most things are about 50% heritable)? Conscientiousness. So the discovery that parents who adhere to parenting advice have children who adhere to school rules is absolutely worthless until you control for conscientiousness - after which the finding should disappear.

To take another example, studies frequently find that parents with a loving, supportive relationship with their children tend to raise happy and cooperative children, and parents with a confrontational relationship with their children tend to raise bratty, defiant children. Harris turns this on its head and says: if a child is happy and cooperative, parents will probably develop a loving and supportive relationship with them. If a child is bratty and defiant, parents will probably develop a confrontational relationship with them. This is sufficiently obvious that any study that just correlates personality and style will, again, be absolutely worthless. Figure out some way to control for this correlation and the connection between parenting style and personality again should disappear.

Harris thinks that these sorts of problem explain the much-trumpeted findings that kids from single-parent homes and children of divorce tend to turn out worse. After all, what kind of fathers abandon their partners and young children? Low conscientiousness fathers who probably have a lot of personal issues. So what kind of children would we expect them to have, just by genetics alone? Low conscientiousness children who probably have a lot of personal issues. And surprise! Children of single parent homes are low conscientiousness and have lots of personal issues! But - and here's something I had never read before - this is true only of homes that are single parent because the father left. If the father died - in a car accident, of cancer, whatever - those children turn out exactly as well as children of double-parent homes! Exactly what one would expect if the problem were caused by what the split implied about genetics and social situation rather than by the parenting itself.

It's not surprising that children don't model behavior they learn from their parents. Parents are horrible people to learn from. First of all, their role in society is completely different from that of children - if a kid sees her parent driving a car, or arguing with a teacher, that's something the kid shouldn't copy - but much of parent behavior, maybe a majority, is like that. Second, parents' interactions with their children are completely uncharacteristic of any other interaction they should expect to encounter; imagine learning social politics from a parent who ends all her arguments with "because I said so", or social norms from a parent who lets her kid get away with things because she's "so cute".

Instead, Harris thinks that children are mostly socialized by

other children. That's why children want, let's say, baseball cards and Pokemon even if their parents collect stamps; more importantly, it's why immigrant children usually grow up speaking most naturally and fluidly the language that they learn in their peer groups rather than the language they learn at home. She backs this up with anthropology, primatology, and evolutionary psychology - in most hunter-gatherer tribes, most chimp bands, and most societies before the Industrial Revolution, parents pretty much just threw their children at the other children in the tribe after age three or so and didn't interact with them much besides feeding them and giving them a place to sleep. The children spent most of their time in mixed-age playgroups that did most of the heavy lifting of socializing them.

In fact, until about 1900, this idea that parents were responsible for raising their children didn't really exist. This bothers me. At this point it's easy for me to believe that things we take for granted in our society are culturally conditioned and may not be true for some godforsaken tribe in the mountains of New Guinea, but to have them be younger than my great-grandmother and *still* have me think they're the natural state of the human condition is pretty atrocious. I guess all those conservative bloggers are right when they say you've got to read old books or else you won't even realize how trapped in a modern worldview you are.

I'm pretty convinced by her arguments. Which is too bad, because it means our society is expending crazy amounts of effort in completely useless directions. And it also raises some bigger problems. For example, if about a hundred years worth of scientists have been wrong about something as big and as obvious as "Parenting style influences your kids'

personalities", then *what else is science wrong about?*

Take the idea of "major calibration failures". That is, right now I think there's practically no chance that Bigfoot or the yeti exists. But if it were discovered Bigfoot really did exist, then instead of saying "Okay, you were right about Bigfoot, but *obviously* there's no yeti, that's just crazy", I would have to say "Wow, whatever thought processes I was using for cryptozoology seem to have been completely flawed; for all I know there might be yeti too. Or a Loch Ness monster."

If I were to learn ghosts really existed, that would be even worse - I could at least admit Bigfoot without accepting that the entire physicalist worldview was wrong. If ghosts turned out to exist, I would have to pretty much re-evaluate *everything* - numerology, reincarnation, God, demons - all would become relatively plausible.

So the bigger a deal I admit I was wrong about, the more I have to accept I might be wrong on a greater number of similar matters. I don't think "parents have no effect on their children's personalities" is as big a deal as "ghosts exist", but it does make me worry how much of (social) science is total bunk.

On the other hand, it's also encouraging. The typical view of scientific controversies is still pretty Galilean: there's this believe that some iconoclast points out that the orthodox establishment is wrong, and then the orthodox establishment spends the next few decades trying to grind them into dust and condemning them as stupid and evil, and their view only comes to be accepted after all the orthodox leaders are dead and a new generation has taken over. That doesn't seem to be

what's happening here.

Judith Rich Harris wrote her book from a position mostly outside the field, most of the orthodox developmental psychologists shrugged and said "Huh, we never really thought about that", and although certainly not everyone has come around to her point of view her theories are being discussed widely and respectfully in the community and a new generation of students is already being taught that this is an interesting controversy. She gets her articles published in mainstream journals and apparently won some big prize for best new psychology research.

So although it doesn't look good for scientists' intelligence not to have come up with these sort of critiques before, it seems relatively complimentary to scientists' integrity and open-mindedness. And (I hope) it doesn't necessarily touch hot-button issues like climate change scientists vs. climate change deniers, or academic medicine vs. alternative medicine, because those are all situations where scientists know that people disagree with them, have read the arguments against them, but still continue believing they're right and the other side is stupid.

Overall I've raised my probability that there are important flaws with modern scientific paradigms that no one has really brought up, but decreased my probability that any particular "heretical" community that says a specific science is flawed is correct.

The Death of Wages is Sin

Federico gives a 6 point plan to cure youth unemployment. It is less complicated and revolutionary than his usual fare; he suggests policies like abolishing the minimum wage and slashing labor regulations. I expect it would work exactly as well as he thinks it would.

After all, minimum wage cuts the bottom out of the labor market. Everyone who would otherwise be working in jobs worth less than $7.25/hour suddenly becomes unemployed. This seems like a bad thing. People making $6/hour seems better than people not being employed at all, right?

Back when there were communists around, some of them would fight against minimum wage laws or occupational safety regulations. Their theory was that these would dull the pain just enough to make workers hate their bosses less and prevent revolution, but not enough to matter. The medical analogy would be a patient who comes in with bone pain, receives a painkiller that pushes the pain back below the threshold of "annoying enough to make me visit a doctor", and never bothers seeking further medical treatment on what turns out to be bone cancer.

I admire the communists for their sheer Xanatosishness, but I don't know how kindly historical hindsight has treated their strategy. On the one hand, they were dead right that better working conditions dampened interest in communist revolution. On the other hand, it seems relevant that a communist revolution would've been horrible, whereas a series of progressively stronger labor regulations actually achieved far more than the communists would have reasonably expected. So this sort of gambit seems potentially very risky.

But this is how I would question whether people making $6/hour or $3/hour or whatever is obviously better than their not being employed at all.

[Before you continue, read this Mother Jones article (h/t: commenter "Nestor") to calibrate your notions of how bad jobs can be for the rest of the article.]

There are probably a lot of people whose labor just isn't worth $7.25/hour. There are probably a lot of people whose labor just isn't worth $3/hour.

As technology continues to advance, I expect the number of these people to increase. I have been accused of the Luddite fallacy for this and I accept the challenge that the historical data present. But there's also this *reductio ad absurdum* where we can manufacture androids exactly as smart as humans in every way for $1. In this world, it seems obvious that all companies would buy androids (who work for free) and fire all their human workers, meaning an end to human employment.

So what's the difference between the past, when technological advances have never caused long-term unemployment, and the android-world, where it does? My guess is that in the past, there have always been areas to shunt the displaced human workers to: maybe machines can manufacture cars, but they can't drive taxis; maybe machines can sew textiles, but they can't predict fashion trends. Technological employment will become a problem only when machines can do *everything* better than humans, which won't be until after the Singularity, by which time we will have much bigger problems to worry about.

Except that's not really true. There may come a time when machines can do most blue-collar jobs better than humans even if they haven't mastered the white-collar ones yet. And

shunting former blue-collar employees to white-collar jobs seems like a hard problem. I don't even think the hard problem is IQ, I think it's some sort of meta-education which is complicated enough that society hasn't figured out how to train it yet. No doubt some blue-collar people will be able to adapt to white-collar jobs, and other people won't. As tech level rises and we approach the android scenario, the number of people who can't adapt gets larger and larger.

Suppose we do what Federico wants. We promote full unemployment. Well then, these growing masses of people aren't going to be unemployed. They're going to be underemployed at $3/hour or something like that.

The minimum wage is sometimes called "the living wage", and there are both lots of sob stories about how it's impossible to support yourself on the minimum wage, and lots of counter-sob-stories by people who claim it's totally easy as long as you don't blow it all on alcohol and expensive hookers. I don't know enough to have a strong opinion but my guess is it could go either way depending on circumstance. But I *am* pretty sure $3/hour is not a living wage. $3/hour either necessitates you to work 20 hour days, or actively drains your income because having a job is expensive (commuting costs, professional clothing costs) but people refuse to give you charity if there are $3/hour jobs available and you haven't taken them.

On the nationwide scale, which is less dystopian? One in which half the population is unemployed and living off government benefits? Or one in which half the population is working 20 hour days at $3/hour jobs like the ones in that Mother Jones article and still struggling to support themselves?

The former situation seems very likely to evolve into a Basic Income Guarantee, about which I have written before in a very similar context and which seem like a proper end state for the economy which may even be preferable to our current situation in many ways (and of course after a basic income guarantee is in place there'll be a much stronger argument for eliminating labor regulations) But the latter situation seems like a disaster, and worse a *stable* disaster that no one has any incentive to make less disastrous.

This strikes me as the strongest argument for the minimum wage and other job-killing labor regulations: that they are turning otherwise-miserably-employed people into unemployed welfare recipients. "Too many people are unemployed and receiving welfare" seems more like a problem society will actually try to solve than "too many people are miserably employed", and maybe the solution will actually do us some good.

Thank You For Doing Something Ambiguously Between Smoking And Not Smoking

"Funge" is a funny word. It refers to the thing which fungible things are able to do, sort of along the lines of what an extra unit of a good is going to replace. I don't think it's a real word and I've only heard it used by people connected to the Center For Applied Rationality. This is too bad, as it prevents everyone else from understanding, let along generating, important sentences. Like "Be careful what you're funging against."

Maybe this is why so many people are so careless what they're funging against. Consider our recent discussion of the minimum wage. The minimum wage means no one has to work for below minimum wage. Its desirability depends a lot on whether below-minimum-wage funges against above-minimum wage jobs or against unemployment. That is, if we ban 100 below-minimum wage jobs, do we get 100 above-minimum-wage jobs, 100 more unemployed people, or a mixture of both?

This was also part of the thrust of my argument about drone warfare – it's not funging against peace, it's funging against much worse types of warfare. The same piece cited the status quo bias and indeed these two ideas are probably related.

This, combined with a complicated regulatory environment and sheer bad luck, seems like the best explanation for the trials (both metaphorical and literal) of e-cigarettes.

E-cigarettes (the "e" is for electronic) are pseudo-cigarettes that contains nicotine without tobacco. They don't smell bad,

they don't produce secondhand smoke, and they don't cause cancer. They are strictly better than regular cigarettes in every way. The governments of several countries are doing their best to ban them.

The governments' position is that they are a stealth attempt to trick people who have successfully avoided regular cigarettes into smoking anyway. There's some merit to this. Some of them have nice fruity flavors that might appeal to children. And because they don't produce smoke, they're legal to use indoors, where some people might not be allowed to use the real thing. There might be this tiny contingent of non-smokers who were just waiting for a flavorful and indoor-useable way to get the addictive expensive chemical they have no reason to want.

And yes, this would be bad. Nicotine is addictive no matter how you get it. There are some claims – and I don't yet know how seriously to take them – that the other chemicals in tobacco inhibit monoamine oxidase which further perverts dopamine levels and makes cigarette smoking more addictive than nicotine alone would be, but this is different from saying nicotine isn't addictive at all. Even if nicotine has few ill effects – and in fact this seems to be the case – there is a strong economic and convenience-based argument for not getting addicted to it.

I should clarify that "few ill effects" claim. A massive overdose of nicotine can kill you (so can a massive overdose of caffeine, Tylenol, or vitamins). Nicotine is a stimulant which raises your heart rate and blood pressure a bit (so is caffeine). It may increase the risk of diabetes, but it may treat cognitive impairment. Overall, it seems to have a complicated mix of minor bad and minor good effects, about the same as anything else in health. And like everything else in health,

tomorrow three labs will come out with studies proving it causes cancer, and a fourth will come out with a study proving it prevents cancer, and one of them may even be right.

There are some studies that show that e-cigarettes have “toxic additives”, but these seem to be in ridiculously tiny trace amounts, don’t seem to make it into the vapor or the body of the user, and the entire problem could be solved by regulation anyway if anyone had a desire to regulate them. This entire issue struck me as a red herring and I bet you can buy fish at any market in the country with more toxic additives than the worst e-cigarette on the market.

So let’s accept that using e-cigarettes will get you addicted and set you back a lot of money and otherwise be annoying but probably not deadlier than anything else you do on a daily basis. What then?

Well, in that case, it’s worse than not smoking but much much better than smoking. And whether or not their existence is a good thing depends on what they funge against. Do they funge against smoking tobacco or not smoking at all?

I would have liked to get the CDC’s opinion, but their webpage on the issue is missing and from commentary I gather it didn’t have the information I wanted anyway. But I did find this:

> A June 2011 national study conducted under the supervision of ECH Research of Cincinnati, in conjunction with Opinionnaire, surveyed more than 200 smoker households that use electronic tobacco products and found that 99% of e-cigarette users are either current or past users of multiple forms of tobacco. Approximately 70% of survey respondents said they intended to quit smoking before starting e-cigarettes.

I can't find the original or even so much as a description of what "ECH" stands for. A sketchy online survey claims that 70% of e-cigarette smokers were former smokers.
So let's just say "probably some high number". This seems quite plausible to me. How many non-smokers think "You know, I want a product with all of the addictiveness and expense of cigarettes, but none of the coolness? In fact, I want to look like a chronically uncool recovering addict inexplicably smoking a glowstick."

That same Etter and Bullen paper says that 96% of the ex-smokers said the e-cigarette helped them quit or reduce smoking, and 79% felt they might relapse to smoking again if they didn't have e-cigarettes. Randomized trials seem to confirm this result, with the average smoker in the trial dropping from 19 cigarettes per day to 2 cigarettes per day after trying e-cigs.

It's not really surprising that e-cigarettes work. My current model of cigarette addiction is that it consists of the interaction between (1) nicotine, (2) smoking-associated behaviors which have become associated with the rush from nicotine over time and might have more complicated components like "oral fixation", and maybe (3) a contribution from MAOIs in the tobacco. Normal cigarettes have all three. E-cigarettes have (1) and (2). Nicotine patches have (1) only. Therefore, e-cigarettes should be more useful in quitting than nicotine patches, albeit not a perfect replacement for regular cigarettes. This does indeed seem to be what has been observed.

Some people argue that the effects of e-cigarettes haven't been perfectly studied, that they might be unsafe in some unclear way. And that as a smoking-cessation device, they're technically a medical device and therefore need to undergo as

much study and regulation as any other medical device before being given to the public.

I know there's constantly a debate between the people who want to evaluate each new medical intervention for safetly and the people who want to use exciting new potentially life-saving technologies *now*, and I know that sometimes the former group do turn out to be right. Varenicline is a popular antismoking drug that was eventually discovered to drive people insane in various way and sometimes lead to suicide; although it is still used on people with both an extreme desire to quit smoking and impressive mental fortitude, it's nice that people paid careful attention to the side effects and didn't just give it out like candy.

But e-cigarettes are literally *the exact same thing as something that's given out to anyone who asks in convenience stores, except without the cancer*. To suddenly hold them to an extremely high standard of safety seems like a fallacy of fungibility.

The worst are the people – one of whom has so far appeared in every article I have read on the subject – who say that we should be careful because "Big Tobacco" is pushing them as a "solution" to the problem of declining cigarette sales. First of all this is just factually wrong; most e-cigarettes are made by alternative companies in direct competition with Big Tobacco. Second, if your reasoning strategy is identifying the Evil People and then minimizing their utility, you probably shouldn't be making public policy.

So I'm against banning e-cigarettes, and I'm even against things like taxing them or prohibiting their use in public places, on the grounds that the more smokers are encouraged to switch to e-cigarettes, the better. Like, if a public e-cigarette

ban reduces the number of smokers who switch to e-cigarettes by 2%, you've just killed an extra 9000 people per year – about three 9-11 attacks, or twice the number of US soldiers who died in the Iraq War.

(this is why public health is about a hundred times more important than any other political issue, and even *tiny little marginal issues* in public health are more important to get right than, say, anything you will see people changing their profile pictures about on Facebook.)

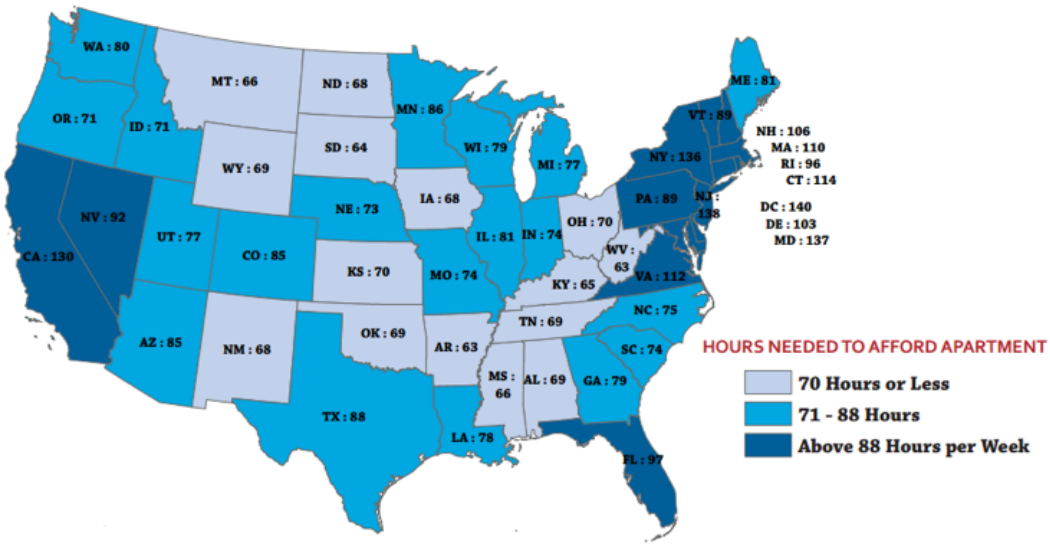
So obvious conclusion is obvious and almost too boring to discuss. I got interested in e-cigarettes because a friend asked me whether he should start taking them *even though he didn't smoke* as a way to get the cognitive enhancement effects of nicotine.

I guess that depends on what you value. I personally wouldn't do this because I'm terrified of addiction. Then again, I avoid coffee and I drink like three glasses of alcohol per year in order to avoid addiction, which most people would consider sort of excessive. And I frequently take weird psychoactive Mexican herbs in order to achieve lucid dreams and would use LSD in a heartbeat if it were legal. So I guess my answer is that my feelings on the costs vs. benefits of various substances aren't likely to generalize across the population.

Lies, Damned Lies, and Facebook (Part 1 of ∞)

I spend so much time arguing with people about the graphics they post on Facebook that I figured I should at least make a blog post out of it so I can pretend it's productive.

Here's an image I got from a few places earlier this week. The title was something like "Hours Working Per Week At Minimum Wage Needed To Afford A Two-Bedroom Apartment In Different States", and it was usually associated with some text about how it was outrageous that the minimum wage isn't nearly enough to live on:



At first glance – and this is how everyone I talked to interpreted it – this seems to support the "minimum wage is unliveable" hypothesis in a big way. In my home state of California, for example, it looks like a minimum wage earner needs to work 130 hours a week just to be able to pay for a place to live. That translates into working nineteen hour days seven days a week. And if you need that ridiculous schedule *just to pay rent*, it doesn't seem to leave a lot of room left over for anything else.

California might be an unfair example because it's particularly high. The state I'm visiting right now, Utah, is a bit more representative, at

77 hours/week. But even this requires eleven hours a day, seven days a week. So people's outrage seems justified.

What's the catch?

The first catch is that this has nothing to do with the minimum wage. The federal minimum wage sets an effective floor for state minimum wage; many states exceed the federal number but none are allowed to go below it. The states that have the best minimum wage apartment affordability – Arkansas, South Dakota, and West Virginia – all have this minimum permissible minimum wage. On the other hand, the states with the highest minimum wage in the country – Vermont, Washington, and Nevada – have exceptionally poor apartment affordability. I don't have numbers to put into SPSS, but a very quick eyeballing suggests apartment affordability as measured by this map and minimum wage are actually *anti*-correlated, probably because high minimum wage implies leftist politics implies dense urban population implies costly housing.

The second catch is related to the first catch: there is no way raising minimum wage could solve this problem. For example, suppose we decided that it was unfair to make people work more than the standard forty hour workweek. What level could we set minimum wage in order to achieve this noble-sounding goal? In California, we would have to multiply the minimum wage by 3.25x, making it $26/hour. I think even most leftists would start to worry that might cause certain problems down the line.

The third and most important catch is that these numbers don't mean what you think they mean and probably don't mean anything at all

I first noticed this when I tried to calculate the price of an apartment in California from this data. California minimum wage is $8/hour, so a quick $8/hour * 130 hour weeks needed * 4 weeks/month = average apartment in California costs $4160/month. Renting apartments in California is horrible, but not *that* horrible.

So I Googled this until I finally found an article in the New York Times where this image had previously appeared, which unlike

every instantiation of the image I have seen on Facebook explains the methodology. The numbers are not about how many hours are needed to *pay for* rent, they're about how many hours are needed to *afford* rent, where afford is arbitrarily defined as "be able to pay for rent using no more than 30% of your income". This brings the cost of a California apartment *in the sense everyone is naturally interpreting the picture as meaning* back to a slightly-less-unreasonable $1250/month and brings the number of hours per week our hypothetical minimum wage laborer needs to work to *pay for* rent down to 39 – still worrying, but markedly less so (her Utah counterpart only needs 23.1, which actually sounds doable and liveable).

But we're still not done here. The graphic specifies that we're affording a two-bedroom apartment. The best reason to demand a two bedroom apartment is that you are, in fact, two people. Suppose we're talking about a couple, both of whom earn minimum wage. Now each partner in California only has to work 19.5 hours per week. Each partner in Utah only needs 11.5. This is starting to sound pretty good.

Can we bring these numbers down further? We might note that even if the average California apartment requires 19.5 hours/week, a minimum wage earner might not be going for the average apartment. They might be after something more modest. How much does a modest apartment cost?

I briefly wondered whether the Internet would be able to tell me which Californian county had the exact median land value for California. Then I remembered this is the 21st century and went straight to Wikipedia's list of California locations by per capita income, which ought to track land value well enough. The exact median is Amador, but since I don't know where that is I chose Sacramento, which is just next to the median, as my experiment. I asked http://sacramento.apartmenthomeliving.com/ to tell me the rent of two bedroom apartments in Sacramento. Its "sort by price" feature is hopelessly buggy, so I wasn't able to find an exact median,

but I am prepared to believe it is about the $1040 the graphic suggests. And yet it is easy enough to find decent 2 bedroom apartments for as little as $650. If we pretend this case is typical, we can declare that a cheap (but liveable) apartment costs about 62.5% of an average apartment. That means each partner in the Californian couple only has to work 12.2 hours/week, and each partner in Utah only has to work 7.2 hours/week.

In fact, these are *still* overestimates for a bunch of reasons. Minimum wage earners are probably concentrated disproportionately in poorer areas of a state, so taking the exact median area of a state overestimates the affordability challenges minimum wage earners face. Most couples share a bed, so they're probably not after a two bedroom apartment and can look for a cheaper one bedroom apartment. And if they really need to, they can do what my roomates do, which is move into a larger house with more rooms and split up the rent even more. In my own living arrangement (two bedroom house where one bedroom is occupied by a couple and the other is occupied by a single person), each partner of the couple only has to work half as hard as the partners in the couple above.

But let's ignore these additional factors and conclude with our 12.2 number for California. Note that this is *less than a tenth* of the number on the original graphic, and is probably a heck of a lot closer to what people think when they read what the graphic is trying to do.

I don't mean to trivialize the problems that minimum wage earners go through trying to pay rent, and certainly not the problems they would go through if they don't work full-time, or are supporting a non-working family member. It's just that this image has nothing to do with these problems and its numbers might as well be generated with random dice rolls for all the good they do anyone.

The Life Cycle of Medical Ideas

I.

About five years ago, an Italian surgeon with the unlikely name of Dr. Zamboni posited the theory that multiple sclerosis was caused by blockages in venous return from the brain causing various complicated downstream effects which eventually led to the immune system attacking myelinated cells. The guy was a good surgeon, nothing about the theory contradicted basic laws of biology, and no one else had any better ideas, so lots of people got excited.

As far as I can tell, the medical community responded exactly one hundred percent correctly. They preached caution, urging multiple sclerosis patients not to develop false hope. But at the same time, they quickly launched studies investigating Zamboni's experiments and used newly gathered data to test the theory. All the results that came back made the idea look less and less likely, so that to my understanding by now it is pretty much discredited. Having successfully spent hundreds of thousands of dollars to empirically disconfirm Zamboni's hypothesis, we can now reflect at leisure on the reasons it was kind of dumb and we should have realized it all along.

II.

About five years ago, two Israel doctors named Gat and Goren posit the theory that benign prostatic hyperplasia, a prostate disease that affects millions of older men, is caused by incompetence of the spermatic veins. They claim they can treat it surgically, and show off rows of smiling patients with glowing testimonials. Once again, the guys are good doctors,

nothing about their theory contradicts basic laws of biology, and no one else has any better ideas.

I shamefacedly admit I *want* this one to be true. There's so much "well, everything is a complicated combination of genes, biomolecules, biopsychosocial stressors and immune modulators that we may never really understand" going on in medicine today that it would be super gratifying if this one mysterious disease turned out to just be plumbing going in the wrong direction. And although the prostate is about as far from my area of expertise as it is possible to be, I have to say that from a physiological standpoint their theory seems to have that rare and much-sought scientific elegance, where everything comes together in a pretty package.

On the other hand, it sounds a whole lot like the Zamboni debacle transposed to a different organ, and Gat and Goren don't have much evidence other than a pretty theory and their own anecdotal success.

As far as I can tell, the medical community has totally ignored this one. Gat and Goren have published their hypothesis and their apparent excellent results in peer-reviewed medical journals. It has garnered praise from prestigious figures in the field (bonus points for calling it "seminal", especially if the pun was intentional). As far as I know, no one has attacked it or even formally expressed doubt. Yet as far as I know, it has gone nowhere.

Does everyone mutually assume that if something this revolutionary were true, someone would have noticed beyond a single article in a urology journal? Do they just decide it needs further research, and hope that this research will be conducted by someone else? Or do they think that it would end up like Zamboni's MS cure, with hundreds of thousands of

dollars wasted, dozens of unnecessary surgeries performed, and nothing to show but yet another fringe medical idea that sounded good at the time?

III.

Minocycline is a relatively boring umpteenth-line antibiotic sometimes used to treat acne. About five years ago, some Japanese doctors noticed that their schizophrenic patients with acne seemed to be *getting better*. This was especially bizarre because some of these patients had "negative symptoms", a set of schizophrenia symptoms considered totally untreatable and which the super-advanced next-generation antipsychotics being pumped out by drug companies can't even touch. They started wondering – can minocycline, an uninspiring antibiotic from the early 1970s, do what all of these psychiatric medications can't?

Once again the medical community responded correctly. They launched a couple of double-blind placebo-controlled studies of minocycline, and sure enough, the stuff was shown to work again and again.

And yet the psychiatrists I know have never heard of it, and I am not aware of any psychiatric hospital in the world where minocycline is routinely given to schizophrenic patients with negative symptoms outside of a clinical trial.

It's not like this is some kind of experimental drug that might kill the patient and isn't even legal yet and we have to wait for further research. If the schizophrenic patient happens to get *acne*, the psychiatrist will be perfectly happy to send them to the nearest CVS Pharmacy to pick up a bottle of minocycline, which they will no doubt have in droves. It's just the schizophrenia connection that isn't there.

I'm totally in favor of waiting for all the research to come in and not jumping to conclusions. The problem is that I don't understand exactly what the process is. If the rule was "We must wait for NIMH to fund a study with greater than 2,000 subjects, and after that everyone will prescribe it, and NIMH is currently working on crunching the data, so just hold your horses," this would sound totally reasonable to me. The problem is that I don't know what we're waiting for and I'm not sure there actually *is* a thing we're waiting for except a spontaneous change in the zeitgeist, which could take forever.

IV.

When people blame drug companies for suppress any promising medications they can't make a profit off of, those people are missing the point.

The drug companies don't suppress promising medications. Promising medications start off pre-suppressed. In some cases they are suppressed by regulation that says a drug has to go through crazy expensive trials before it can be approved. In other cases, they are suppressed simply by the burden of proof: even without the government, doctors aren't going to prescribe something they don't know is safe and effective, and they're not going to know it's safe and effective without studies, which as I may have mentioned are crazy expensive. In still other cases, the medications are suppressed by medical conservativism: most doctors very reasonably don't want to use a drug unless they know other doctors they respect are using the drug, so unless the drug impresses itself onto the consciousness of the entire medical community at once it will fizzle out.

What drug companies do, as best I understand it, is put billions of dollars and millions of man-hours of effort into un-

suppressing those particular drugs it is in their financial interest to un-suppress. They are doing a great service. It's just a very selective one.

I'm not sure how it works in surgery. As far as I know, there aren't companies that patent surgical procedures and then popularize them. If there were, maybe one of them would pick up Gat-Goren and give it a fair try to stand or fall on its own merits. As it is, it looks like it will have to wait for some university or charitable group to pick it up – and let's face it, "my eighty year old grandpa gets up to piss half a dozen times a night" isn't quite as sexy as multiple sclerosis.

In medicine, drugs are usually approved for specific indications. Doctors are allowed to prescribe them for other indications, but there are trivial inconveniences and minor legal hurdles and in practice most of them rarely do. Some pharmaceutical company was nice enough to get minocycline approved for acne back in the '70s, but since then it's gone off patent and no one owns it enough to say "Hey, start the process to approve this drug of mine for schizophrenia!" The medical community is pretty smart and I bet there's a process by which this will eventually happen, but I also bet it will take a long time and be overly complicated and a whole lot of schizophrenics will have to suffer from negative symptoms long after the vanguard of the medical community has satisfied itself that these are treatable.

(I can imagine the look on my attending's face if I suggested we treat one of our schizophrenic patients with minocycline. I expect I would get a lecture on how We Have To Be Responsible And Ethical, and then we would give them one of the same three drugs we give all schizophrenics. I might have more luck painting little fake acne pustules on my schizophrenic patients' faces, but most of the ones I have now

are Paranoid-Type and I *really* can't imagine them going along with that. I'll just have to wait until I get someone catatonic.)

Which is why it sucks that the *other* really interesting drug that might revolutionize the treatment of schizophrenia is an antihypertensive from the 1950s.

V.

I find the life cycle of medical ideas really interesting.

I was always taught that there were two kinds of medicine. Real medicine, which has been proven to work by studies. And alternative medicine, which has been proven not to work by studies but people still use it anyway because they are stupid.

This dichotomy leaves out the huge grey area of "things that seem like they will probably work, and a few smaller studies have shown very promising results, but no one has bothered to do larger studies, or if they have they have never really been incorporated into medical practice for reasons I can't put my finger on."

Some of this grey medicine, like Zamboni's MS treatment, are doomed to eventually fall back into the abyss of alternative medicine. Others, like the Gat-Goren procedure, teeter in the middle, threatening to go either way. Still others, like minocycline, have already been sanctified and dressed in robes of white, and the only thing preventing them from entering Evidence Based Heaven is some sort of weird bureaucratic snafu at the Pearly Gates.

I am encouraged that all three of the examples of grey medicine cited in this article are about five years old. It suggests that there's a certain window of time during which grey medicine is well-known but hasn't yet been well-studied.

Maybe most of the newer stuff I don't know about, and most of the older stuff has been successfully proven or disproven. It seems possible to me that the current system does have the optimal combination of safety and innovation, or at least the best we can do without a Science Czar. As I immerse myself in Medical Culture, I look forward to finding out whether there are some hidden processes for dealing with this, or whether the situation is really as dire as it looks.

But I am also hopeful that some new organizations like Microryza and MetaMed might, totally independent of justified or unjustified medical conservativism, be able to speed the process along.

Vote on Values, Outsource Beliefs

I.

Today I learned about social impact bonds. They are a thing that exists. I would expect them to be in an adequate civilization like Raikoth or dath ilan. But they are a thing that exists on Earth.

The basic idea is: government could save a lot of money if some problem got fixed. For example, if people stopped committing crime, they could spend less money on prisons. So they make a deal with a corporation. The corporation agrees to spend a certain amount of money to prevent crime for five years. And if crime goes down and the government saves on prisons, the corporation gets half the savings (or a third, or whatever).

Zero taxpayer money gets risked. It is entirely up to the corporation to fund the problem-solving effort. If they fail, then it's their own loss. If they succeed, then the government pays them money, but less than the government made, so the taxpayers still get a profit.

(The main exception I can think of is if by coincidence, crime was about to drop by 50% anyway right when the program started, and the government ends up giving half of its prison savings to the corporation for no reason. But presumably you hire a couple of mediocre economists and they are able to price out this risk. Also, a lot of the social impact bonds use a slightly different method of assessment, where they compare crime among the people the corporation has helped to crime among a control population to be sure it was the intervention that did it.)

The particular article I read about this today was How Goldman Sachs Can Get Paid To Keep People Out Of Jail. It was the name "Goldman Sachs" that got me excited. They're an investment bank. Their job is predicting risk. I don't know if they're any good at it or not. But they're the sort of organization that potentially could be. So we have people who understand risk trying to figure out what social policies will produce which results, with money riding on the decision.

This is looking *impressively* close to prediction markets. Futarchy says "vote on values, bet on beliefs". Asking a corporation to invest money in crime-solving is a form of betting on belief – they are betting on what anti-crime programs will decrease crime most and win them the most reward. You still have the elected government deciding what bonds to place – voting on values – but you're outsourcing your beliefs to the corporation involved and giving them an incentive to get it right.

Think of all the possibilities.

Right now we have a system where we don't really help people in need, unless the need becomes desperate, in which case we would feel bad about not helping, so we do, but then the cost of helping has gone up by an order of magnitude. This is exactly the sort of stupid thing that a market should be able to profit from solving.

We could have a health insurance company giving free preventative care to the poor, and the government paying them out of decreased emergency room visits.

A psychiatry clinic giving therapy to at-risk patients, and the government paying them out of decreased involuntary commitments.

A university accepting students without tuition, and the government paying them out of the increased tax revenue when they take higher-paying jobs.

Planned Parenthood offering free IUDs for women who need them, and the government paying them the money it saves from not having to put the kids through school.

Trade schools offering free classes to people on welfare, and the government paying them back from not having to give them welfare checks once they get good jobs.

I'm not sure what it means that we're not doing those sorts of things already. But if we can't figure out a way to solve those problems without bringing in a corporation to profit off of our incompetence, I say bring in the corporations.

II.

I think many people are against government social programs for a lot of the same reason that *The Last Psychiatrist* is against maintenance of certification exams (a position I totally called). There's too much temptation to use it as a signal that you are Doing Something while in fact funding programs like DARE which look virtuous, but do nothing or even actively make the problem worse.

If you lean this way – and I think I do – then it is not solely out of stupidity that we wait until problems have become dire before doing anything about them. Yes, it would be great to give free job training to people on welfare and save money when they come off welfare more quickly. But actual job training programs for welfare recipients are abysmal and have been denounced as a "charade" from both the left and the right. They may be a lost cause, but I would like to see someone who has an incentive to succeed try first before

writing them off – or at least get the evidence that would be provided by no such person being willing to try.

III.

For a while I was confused by the old libertarian talking point that "greed is good". I think I could phrase it a little better now. Greed isn't *good*, per se. It is *honest*. You know where you stand with greed. You never wonder if greed has an ulterior motive, because it's already the most ulterior motive there is. Greed feels no temptation to corruption, because the thing it would do if it were corrupt is precisely what it's doing anyway. Greed is like the Harlot in one of Khayyam's rubaiyat:

> A Sheikh beheld a Harlot, and said he:
> "You seem a slave to drink and lechery"
> Replied the Harlot: "What I seem … I am!
> Are you, O Sheikh, all that **you** seem to be?"

As I see it, capitalism isn't about worshipping greed, but about figuring out how to make greed work for good ends. So far, it has mostly tried to apply greed to get us cheap and attractive consumer products. And the amount of cheap and attractive consumer products is, like, the one thing that everyone can unambiguously agree our civilization hasn't dropped the ball on. If we all die tomorrow and aliens discover Earth ten thousand years from now, their anthropologists will publish books saying "They sure were screwed up, but *man* did they have a lot of cheap and attractive consumer products."

And I think some of the most exciting proposals for the future involve finding ways to use this privileged incorruptible perfectly-incentivized status of greed to do other things. Prediction markets are promising because they use greed to fix

epistemology. Neocameralism is promising because it uses greed to fix governance. And social impact bonds are promising because they use greed to fix social problems.

…which isn't to say it's going to be easy. Ozy's first response is that Goldman Sachs should use their $10 million to give ten thousand people in the control group a $1000 bribe each to commit a small crime; this will be more than enough to demonstrate a *vastly* reduced probability of criminality by being in the intervention group and earn Goldman $20 million.

I told Ozy zir plan is unnecessarily complex. Look at the numbers. Two hundred potential criminals. And they need a 50% decrease in jail time to meet their target and earn $20 million.

So go to the potential criminals and tell them "I'll give you $50,000 to not commit any crimes in the next few years. $25,000 now, in order to help you solve whatever problems turned you to criminality. And $25,000 at the end, after you've successfully avoided jail, as a reward." If half of them stick to it, then boom, you get $20 million and you've made a $10 million profit. And incentivized the next generation of criminals, but you've already *got* your profit, that's the next generation's problem.

The fact that this would *work* probably says a lot about the inefficiency of prison compared to any other conceivable way of dealing with crime. And about the profits Goldman Sachs or anyone else willing to face the inefficiency head on could make.

I don't know if it's exactly a *good* idea to bring in the people who caused the financial crash to help the people who came up with the prison system. But since all we've got is incompetent

institutions, maybe sticking *different* incompetent institutions in different roles might at least shake things up a little.

A Something Sort of Like Left-Libertarian-ist Manifesto

I.

> "Forgive him , for he believes that the customs of his tribe are the laws of nature." –**George Bernard Shaw**

Our tribe has a custom of dividing into Right and Left. The Right supports economic laissez-faire and traditional social norms. The Left wants economic regulations and greater civil liberties.

(unless of course a Democrat is in office)

If you live too long under this system, you start thinking the Left-Right division is a law of nature. I like the Libertarians' pet Two Dimensional Political Compass because it reminds people that they're allowed to mix and match.

And so, glory be unto the infinite variety of human thought, we have moved from an unwillingness to credit more than two possible visions of a flourishing society to a grudging acceptance that maybe there are as many as *four* such visions.

(one of which nobody will admit to believing)

"The limits of our language are the limits of our world". If the only two words in political discourse are Left and Right, it becomes hard to realize libertarianism is a possibility, let alone evaluate it. What equally coherent possible views might a four-word discourse be missing?

What if we abandon our tribe's custom of conflating free market values and unconcern about social welfare?

Right now some people label themselves “capitalists”. They support free markets and oppose the social safety net. Other people call themselves “socialists”. They oppose free markets and support the social safety net. But there are two more possibilities to fill in there.

Some people might oppose both free markets and a social safety net. I don’t know if there’s a name for this philosophy, but it sounds kind of like fascism – government-controlled corporations running the economy for the good of the strong.

Others might support both free markets and a social safety net. You could call them “welfare capitalists”. I ran a Google search and some of them seem to call themselves “bleeding heart libertarians“. I would call them “correct”.

II.

I think I only realized how committed to this position I was when I read an article about the BART strike. Workers on the BART, a San Francisco area mass transit system, were striking for higher pay. A tech CEO suggested solving the problem by firing the workers and automating their jobs. Some other people didn’t like that, said that BART Worker was one of the only jobs that people without college education could get and make good $60,000+ salaries, said employees were mostly old and wouldn’t be able to get other work, said even if their jobs could be automated it would be cruel to destroy their livelihoods just for the sake of profit.

And my first thought was: if your job can be done more cheaply without you, and the only reason you have it is because people would feel sorry for you if you didn’t, so the government forces your company to keep you on – well then, it’s not a job. It’s a welfare program that requires you to work 9 to 5 before seeing your welfare check.

Suppose BART work really can be done just as well by a cheap machine. Compare the current system – in which BART is prohibited from firing the workers and replacing them with the machine because that would be greedy – to a system where BART fired the workers, bought the machines, but continued giving the workers their old paychecks for no reason. BART gets the same profits either way. The workers get the same amount of money either way. The only difference is that the workers gain forty hours of free time a week.

That suggests that long hours worked by BART employees under the current system are deadweight loss, and the role of BART work is the same as those legendary New Deal welfare programs where they made people dig ditches and fill them in again.

Assuming society has decided it wants to give people welfare, it can do it in one of two different ways: the traditional way, where the government sends them a simple welfare check once a month. Or the sneaky way, where it gets billed as a "job" at the BART.

In the "Simple Check" condition the welfare is funded by the tax base, which presumably is the general population, with rich people paying significantly more. In the "Sneaky Job" condition, the welfare is funded by mass transit users – disproportionately poor people – and the increased cost inevitably disincentivizes mass transit. You may remember mass transit as the thing that cuts down on traffic, sprawl, and carbon emissions – you know, that thing we are trying to desperately convince people to do more of.

In the "Simple Check" condition the recipients of the welfare are the entire impoverished population, although the system may place more emphasis on those who are poorer or need

more. In the "Sneaky Job" condition, the recipients of the the welfare are those few well-connected people who get cushy jobs at the BART, chosen somewhat at random but with the usual biases of employers being more likely to hire attractive, tall, Caucasian, etc people. They get $60,000 + no doubt excellent benefits, and everyone else misses out.

In the "Simple Check" condition, the recipients of the welfare can live enjoyable lives doing their hobbies – as the woman in the article puts it, hair and makeup. In the "Sneaky Job" condition, the recipients have to work long hours doing busy work, suffer the normal vagaries of jerkwad bosses and office politics, and suffer the constant stress that they might be fired for underperforming.

With all these advantages of "Simple Check", what exactly is the "Sneaky Job" condition good for that makes it so popular? As far as I can tell, it is good for fooling people. People do not like paying welfare. But if welfare is placed in work boots and wears a big sign with the word "JOB" painted on it in bright letters, they will walk by it without grumbling. Also important, people do not like *being* on welfare, and as the Rogers & Hammerstein song goes, "when I fool the people I fear, I fool myself as well".

[lest I be accused of being insensitive by pointing out how *other people's* jobs are welfare, I will freely admit I have a job partly because the government pays my hospital $100,000 to employ me (of which I get less than half). This is a sufficiently complicated system that a full explanation will have to await another post.]

III.

Welfare has even more clever disguises than this. Let's talk about those fast food workers who want $15 an hour.

No one denies that it's pretty crappy to have to live on $8 or so an hour, which is about what fast food workers currently make. But if fast food workers get $15 not because they do $15 worth of work, but because we feel sad that they're living on too little money, then once again it's welfare.

And once again we can give them that welfare in one of two ways. We can send them a check, or we can pressure fast food places to pay them more.

If we send people a check, it goes to everyone, whether employed or unemployed. If we pressure fast food places to pay more, then it's only employed people – the people who need money the least – who get anything.

If we send people a check, who gets the check is presumably determined by need. If we pressure fast food places to pay more, then who pays more is determined by media exposure and political clout. Fast food workers seem to have good union and good public visibility, so they can demand their wages get raised to $15. Garment workers aren't as well-organized or are less sympathetic, so their wages stay at $8. It encourages a system of "squeaky wheel gets the grease" in which "squeaky" means "go on strike a lot and act miserable".

If we send people a check, the costs are passed on to the taxpaying public, which includes rich people who pay extra taxes and does not include poor people who get out of a lot of taxes because of their low income. If we pressure fast food places to pay more, the costs are passed on to fast food consumers, who are less likely to be wealthy and more likely to be black than the general population.

And if we send people a check, there's not much taxpayers can do to get out of the extra cost. But if we pressure fast food companies to pay people more, we punish them for hiring

workers. If the workers do $8 worth of work for the company, and the government makes them pay $15, it's the equivalent of fining companies $7 an hour for hiring poor people. Not only is this morally unfair, but companies will probably respond rationally by automating as much work as they can, hiring fewer people, or trying to figure out how to replace multiple poor people with fewer wealthier people (for example replacing several clerks with a programmer who runs a computer system).

This is a somewhat harder case as the demand for higher wages among fast-food employees seems endogenous – they're threatening to strike and show the companies how much they need them – rather than exogenous – motivated by government fiat or popular demand. Labor negotiations are coordination problems that are more opaque to analysis than I like. But I think a case can certainly be made that here, too, people are shooting for a noticeably inferior solution just because it helps them avoid *thinking* about the poor. It's not about complicated problems or a changing economic landscape – just make that greedy Walmart behave and somehow I will be freed of all responsibility and all consequences.

At the moment, I *might* support higher minimum wages just because doing things the right way is politically impossible. One can make all sorts of stupid political policies attractive when they are combined with other stupid political policies. But I am not pleased about it and any time people say we need minimum wages to "punish greedy corporations" it just makes me question the life choices that have made me end out on the same side of a political issue as they did.

IV.

But combining market values and compassion isn't just about solving everything with basic income guarantees. Let me give another example of a government program meant to increase social welfare and how a more market-informed version would be better than a brute-force regulation.

Affirmative action and minority rights. I don't trust people on this blog to think clearly about any actual minority group, so let's pretend we're worried about affirmative action for Martians, who have been a disempowered underclass ever since their giant heat-ray-bearing tripod machines broke down.

Modern affirmative action says that given the choice between a Martian or an equally qualified Earthling, one must hire the Martian. One big obvious problem here is that "equally qualified" is a matter of opinion. It may be that a boss is prejudiced against Martians, and so tells an excellent Martian candidate that ve is underqualified for the position – the Martian may never know. Or a Martian who was genuinely underqualified may paranoidly believe ve was denied out of prejudice and start a costly lawsuit.

There are other problems as well. Some jobs may have legitimate reasons not to hire Martians – maybe Martians make lousy pilots because their single lidless eye gives them terrible depth perception. Certainly a Martian actor is unqualified to play Abraham Lincoln in a historical biopic. One could offer to let these jobs apply for exemptions, but this means a costly bureaucratic process, and is likely to end with large companies with good lawyers obtaining the exemptions, small companies with poor lawyers not obtaining the exemptions, and no concern about fairness to Martians in any case.

In the worst possible situation, a non-prejudiced boss may decide not to hire Martians because it would be harder to reprimand or dismiss a Martian when they could threaten to sue the company or start a viral Tumblr post accusing the company of speciesism.

Compare a market-informed solution: run a bunch of controlled studies in which bosses get identical Earthling and Martian resumes, find out exactly how strong the prejudice against Martians is, then levy an appropriate tax on hiring Earthlings (or give a subsidy for hiring Martians). Maybe hiring Earthlings costs 5% extra, which is funnelled into scholarships for impoverished Martian larvae.

Now there's no question of a company wriggling out of their obligation – no matter how stylish their lawyers' hair is, they're going to pay the tax. There's no question of lawsuits – if a company didn't hire a qualified Martian, that's their own business and the Martian community can laugh all the way to the bank. But on a *statistical basis*, we expect companies to be indifferent between hiring Martians or Earthlings.

Any company that has a legitimate reason to not want to hire Martians can just pay the (small) tax. And there's no problem with firing Martians anymore – if you decide to fire the Martian in favor of an Earthling you like more, you're perfectly welcome to do so as long as you don't mind paying a little extra.

If ten years later the social scientists do some studies and find that companies are still more likely to accept Earthling resumes over identical Martian resumes, they can raise the tax until that's no longer the case. If they find that companies are more likely to accept *Martian* resumes now, then prejudice has decreased and the tax can decrease as well.

I think everyone has a lot to like about this proposal. Martians can rest assured that with enough time to tweak the tax level, they will have a provably equal playing field in this area. Non-bigoted Earthlings can rest assured that they're not going to be unfairly accused of bigotry and taken to court by some Martian playing the planetary origin card. And bigoted Earthlings who just really don't like Martians – maybe someone's father was killed by a heat-ray-tripod during the invasion and she's had PTSD every time she sees Martians ever since then – can stand by their "principles" as long as they're willing to pay a little extra.

(This is my answer to Jim's question of "How many cities are you planning to burn, how many women are you planning to have raped with large objects in order to achieve equality of opportunity?", which I honestly have to admit is not a question I ever really considered before reading Jim's blog)

Someone will object that small fees can't eliminate as pervasive a social problem as prejudice, but I'm not so sure. Consider the Islamic Caliphate (7th – 12th century AD). Their modus operandi was to march into a new territory, tell the non-Muslims there that they were *perfectly welcome* to continue to practice their old religion as long as they paid a tax, and if they ever wanted to save those couple shekels or dinars or whatever, they could also convert to Islam – but no pressure. The current religious makeup of the Caliphate territory (Northern Africa and the Middle East through Iran and Pakistan) should be taken as some evidence of the effectiveness of this policy.

V.

In my opinion the biggest advantage of a market-based system for improving social welfare is that it allows more flexibility –

it leaves your options open.

Suppose the government, noticing mercury is toxic and has few good industrial uses, bans the use of mercury in industry.

A month later, some chemist discovers a really really lucrative industrial application for mercury that will make billions of dollars and cut the price of automobiles in half.

Probably this chemist can't single-handedly convince the government to relax its views on mercury. She could consider selling her idea to a really big company like Dow Chemical, who could afford the necessary lobbying. But then she's lost the ability to profit from her own invention, and we've replaced what could have been a nimble startup with yet another Dow product that they'll overprice and destroy a couple of Indian villages producing. And the brilliant scientist becomes a mid-level drone working for morons in suits.

Or maybe it's worse than this. Maybe she goes to Dow, but they don't want to take the time to understand this fringe idea. Or maybe Dow is mildly interested but not interested enough to throw all its lobbyists and lawyers at the problem. Or maybe Dow *does* throw all its lobbyists and lawyers at the problem, but the Sierra Club reasonably believes that this is just another evil company trying to gut vital environmental legislation, and successfully blocks them. Sure, Dow says "This will halve the price of automobiles!", but they probably make grandiose claims about *all* of their products when they're trying to look good in front of the government.

So suppose that instead of banning mercury, the government just places a tax on it. The tax could be the cost of mercury cleanup, it could be enough money to treat and emotionally compensate mercury poisoning sufferers, or it could just fund public health programs that do more good than fighting

mercury ever could. It could be all these things combined plus a little extra. Let's say the tax on mercury is 500%. Every company that has any possible alternative to mercury switches to that alternative. The companies that have no alternative to mercury close down if the benefit of their product to society is less than the cost of the mercury they produce. And the companies that use mercury in a way that net benefits society stay open and subsidize lots of environmental and public health programs.

Now the chemist who discovered the brilliant unexpected use for mercury is able to start her startup – at increased costs, sure, but if it's as lucrative an idea as she thinks she'll be able to get the investment or just swallow the losses. In any case, it's nothing compared to the cost of pushing around an entire government agency. The price of automobiles decreases by half, the taxes are more than enough to clean up the mercury and improve public health, and everyone is happy.

VI.

The problem with banning and regulating things is that it's a blunt instrument. Maybe before the thing was banned someone checked to see whether there was any value in it, but if someone finds value after it was banned, or is a weird edge case who gets value out of it even when most other people don't, then that person is mostly out of luck. Even people operating within regulations have to spend high initial costs in time and money proving that they are complying with the regulations, or get outcompeted by larger companies with better lobbyists who can get one-time exceptions to the regulations.

In short, the effect is to decrease innovation, crack down on nontypical people, discourage startups, hand insurmountable

advantages to large corporations, and turn lawsuits into the correct response to everything.

The problem with not banning and regulating things is that the rivers flow silver with mercury, poor people starve in the streets, and Martians get locked out of legitimate industry and are forced to turn to threatening innocent cities with their heat rays just to get by.

The position there's no good name for – "bleeding heart libertarians" is too long and too full of social justice memes, "left-libertarian" usually means anarchists who haven't thought about anarchy very carefully, and "liberaltarian" is groanworthy – that position seems to be the sweet spot between these two extremes and the political philosophy I'm most comfortable with right now. It consists of dealing with social and economic problems, when possible, through subsidies and taxes which come directly from the government. I think it's likely to be the conclusion of my long engagement with libertarianism (have I mentioned I only engage with philosophies I like?)

Plutocracy Isn't About Money

Two political science articles I read recently have surprisingly dissonant conclusions.

Gilens and Page's study "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens" is very interesting. You may have spotted it in the news media under any of a host of diverse titles:

The New Yorker: Is America An Oligarchy?

BBC: Study: US Is An Oligarchy, Not A Democracy.

RT: Oligarchy, Not Democracy.

Business Insider: Major Study Finds That The US Is An Oligarchy.

And my favorite, Daily Kos: Too Important For Clever Titles: Scientific Study Says We Are An Oligarchy

(the word "oligarchy" appears in the study only once, at the bottom of page six, as a reference to an alternative theory the authors do not endorse)

But RAMPANT MEDIA PLAGIARISM aside, it's not a bad summary. The study tries to determine what factors predict whether or not a policy gets implemented in the United States. They compare popular support to elite support, where "elites" are the wealthiest ten percent, and find that elite support is a stronger predictor. I believe the way they put it is that once you know whether elites support a policy, learning whether or not the general public supports it improves your model's ability to predict whether or not it gets passed only an tiny amount, even though elite opinion and popular opinion are often quite different.

Also recently, Rationalist Conspiracy had a good post on Money Doesn't Matter In Politics. A lot of anecdotes, but also links to some convincing studies, like the one that shows how "in Congressional races where candidates spent about $250K (1990 dollars), every $100K spent got another 0.3% of the vote, a tiny amount."

To Alyssa's list I would add Ansolabehere, Figueiredo and Snyder's: Why Is There So Little Money In Politics?, recently spotted on Marginal Revolution. The summary (which does not include the word "oligarchy"):

> "We show that only one in four studies from the previous literature support the popular notion that contributions buy legislators' votes. We illustrate that when one controls for unobserved constituent and legislator effects, there is little relationship between money and legislator votes. Thus, the question is not why there is so little money in politics, but rather why organized interests give at all."

I call these "dissonant" because the simplest explanation for the Gilens and Page finding is that the economic elite are buying elections. But the Ansolabehere et al result says they couldn't even if they tried. If we take both of these studies at face value, how can we reconcile them?

I can think of a few hypotheses:

1. Legislators vote based on their personal opinions. Most legislators are elite, therefore their opinions correlate with the opinions of other elites.

2. Elites control the media, the universities, et cetera. They affect legislators indirectly, by affecting the entire culture (but how would they do this without influencing commoners?

Maybe this is a subset of [1], in that elites consume elite-produced media?)

3. Legislators would *like* to think they are elite, and so they vote with elite opinion in the hopes of looking cool and getting elites to like them.

4. Money does not buy elections, but legislators think it does, so they try to satisfy the people with the money in order to win elections.

5. Money does not buy elections, but money can fund think tanks and lobbyists who can persuade legislators through non-election-buying means. This doesn't take the form of promising financial support or during elections, it just comes from talking and befriending and advising and convincing them. The studies showing money doesn't affect campaigns miss this effect. Ansolabehere seems to like this one, pointing out that interest groups spend ten times as much as lobbying as on direct campaign contributions. But even here there are economic arguments against. They estimate that one hour of a legislator's time costs $10,000. This is a high number, but if talking to legislators seriously affected legislation it would be an *amazing* steal.

6. Elites vote more and are more politically active in terms of volunteering, letter-writing, etc. Legislators try to cultivate their affection to win elections, but it has nothing to do with money. But this effect doesn't seem strong enough to make up for the small number of elites.

7. The connection between elites and successful policies is a coincidence – not in the sense that the study found a nonsignificant finding, but in the sense that elite opinion and legislative success are both biased in the same direction for different reasons. For example, maybe elites tend to lean

conservative, and the conservative party in government is much better organized and able to push more legislation through. Gallup finds there is not a big difference between elites and commoners in terms of basic party labeling. But this study (which does define "elite" somewhat differently) shows that elites are predictably less supportive of welfare and redistribution programs than commoners are (I am enraged that this study doesn't give good comparative data on social issues). If those programs tend to fail for some reason, that could help produce some of these effects.

Against Tulip Subsidies

I.

Imagine a little kingdom with a quaint custom: when a man likes a woman, he offers her a tulip; if she accepts, they are married shortly thereafter. A couple who marries sans tulip is considered to be living in sin; no other form of proposal is appropriate or accepted.

One day, a Dutch trader comes to the little kingdom. He explains that his homeland *also* has a quaint custom involving tulips: they speculate on them, bidding the price up to stratospheric levels. Why, in the Netherlands, a tulip can go for ten times more than the average worker earns in a year! The trader is pleased to find a new source of bulbs, and offers the people of the kingdom a few guilders per tulip, which they happily accept.

Soon other Dutch traders show up and start a bidding war. The price of tulips goes up, and up, and up; first dozens of guilders, then hundreds. Tulip-growers make a fortune, but everyone else is less pleased. Suitors wishing to give a token of their love find themselves having to invest their entire life savings – with no guarantee that the woman will even say yes! Soon, some of the poorest people are locked out of marriage and family-raising entirely.

Some of the members of Parliament are outraged. Marriage is, they say, a human right, and to see it forcibly denied the poor by foreign speculators is nothing less than an abomination. They demand that the King provide every man enough money to guarantee he can buy a tulip. Some objections are raised: won't it deplete the Treasury? Are we obligated to buy

everyone a beautiful flawless bulb, or just the sickliest, grungiest plant that will technically satisfy the requirements of the ritual? If some man continuously proposes to women who reject him, are we obligated to pay for a new bulb each time, thus subsidizing his stupidity?

The pro-subsidy faction declares that the people asking these question are well-off, and can probably afford tulips of their own, and so from their place of privilege they are trying to raise pointless objections to other people being able to obtain the connubial happiness they themselves enjoy. After the doubters are tarred and feathered and thrown in the river, Parliament votes that the public purse pay for as many tulips as the poor need, whatever the price.

A few years later, another Dutch trader comes to the little kingdom. Everyone asks if he is there to buy tulips, and he says no, the Netherlands' tulip bubble has long since collapsed, and the price is down to a guilder or two. The people of the kingdom are very surprised to hear that, since the price of their own tulips has never stopped going up, and is now in the range of tens of thousands of guilders. Nevertheless, they are glad that, however high tulip prices may be for them, they know the government is always there to help. Sure, the roads are falling apart and the army is going hungry for lack of rations, but at least everyone who wants to marry is able to do so.

Meanwhile, across the river is another little kingdom that had the same tulip-related marriage custom. They also had a crisis when the Dutch merchants started making the prices go up. But they didn't have enough money to afford universal tulip subsidies. It was pretty touch-and-go for a while, and a lot of poor people were very unhappy.

But nowadays they use daffodils to mark engagements, and their economy has never been better.

II.

In America, aspiring doctors do four years of undergrad in whatever area they want (I did Philosophy), then four more years of medical school, for a total of eight years post-high school education. In Ireland, aspiring doctors go straight from high school to medical school and finish after five years.

I've done medicine in both America and Ireland. The doctors in both countries are about equally good. When Irish doctors take the American standardized tests, they usually do pretty well. Ireland is one of the approximately 100% of First World countries that gets better health outcomes than the United States. There's no evidence whatsoever that American doctors gain anything from those three extra years of undergrad. And why would they? Why is having a philosophy degree under my belt supposed to make me any better at medicine?

(I guess I might have acquired a talent for colorectal surgery through long practice pulling things out of my ass, but it hardly seems worth it.)

I'll make another confession. Ireland's medical school is five years as opposed to America's four because the Irish spend their first year teaching the basic sciences – biology, organic chemistry, physics, calculus. When I applied to medical school in Ireland, they offered me an accelerated four year program on the grounds that I had surely gotten all of those in my American undergraduate work. I hadn't. I read some books about them over the summer and did just fine.

Americans take eight years to become doctors. Irishmen can do it in four, and achieve the same result. Each year of higher education at a good school – let's say an Ivy, doctors don't

study at Podunk Community College – costs about $50,000. So American medical students are paying an extra $200,000 for…what?

Remember, a modest amount of the current health care crisis is caused by doctors' crippling level of debt. Socially responsible doctors often consider less lucrative careers helping the needy, right up until the bill comes due from their education and they realize they have to make a lot of money *right now*. We took one look at that problem and said "You know, let's make doctors pay an extra $200,000 for no reason."

And to paraphrase Dirkson, $200,000 here, $200,000 there, and pretty soon it adds up to real money. 20,000 doctors graduate in the United States each year; that means the total yearly cost of requiring doctors to have undergraduate degrees is $4 billion. That's most of the amount of money you'd need to house every homeless person in the country ($10,000 to house one homeless x 600,000 homeless).

I want to be able to say people have noticed the Irish/American discrepancy and are thinking hard about it. I *can* say that. Just not in the way I would like. Many of the elder doctors I talked to in Ireland wanted to switch to the American system. Not because they thought it would give them better doctors. Just because they said it was more fun working with medical students like myself who were older and a little wiser. The Irish medical students were just out of high school and hard to relate to – us foreigners were four years older than that and had one or another undergraduate subject under our belts. One of my attendings said that it was nice having me around because I'd studied Philosophy in college and that gave our team a touch of class. *A touch of class!*

This is why, despite my reservations about libertarianism, it's not-libertarianism that really scares me. Whenever some people without skin in the game are allowed to make decisions for other people, you end up with a bunch of elderly doctors getting together, think "Yeah, things *do* seem a little classier around here if we make people who are not us pay $200,000, make it so," and then there goes the money that should have housed all the homeless people in the country.

But more important, it also destroyed my last shred of hope that the current mania for requiring college degrees for everything had a good reason behind it.

III.

The only reason I'm picking on medicine is that it's so clear. You have your experimental group in the United States, your control group in Ireland, you can see the lack of difference. You can take an American doctor and an Irish doctor, watch them prescribe the same medication in the same situation, and have a visceral feel for "Wait, we just spent $200,000 for no reason."

But it's not just medicine. Let me tell you about my family.

There's my cousin. He wants to be a firefighter. He's wanted to be a firefighter ever since he was young, and he's done volunteer work for his local fire department, who have promised him a job. But in order to get it, he has to go do four years of college. You can't be a firefighter without a college degree. That would be ridiculous. Back in the old days, when people were allowed to become firefighters after getting only thirteen measly years of book learning, I have it on good authority that several major states burnt to the ground.

My mother is a Spanish teacher. After twenty years teaching, with excellent reviews by her students, she pursued a Masters'

in Education because her school was going to pay her more money if she had it. She told me that her professors were incompetent, had never actually taught real students, and spent the entire course pushing whatever was the latest educational fad; however, after paying them thousands of dollars, she got the degree and her school dutifully increased her salary. She is lucky. In several states, teachers are required by law to pursue a Masters' degree to be allowed to continue teaching. Oddly enough, these states have no better student outcomes than states without this requirement, but this does not seem to affect their zeal for this requirement. Even though many rigorous well-controlled studies have found that presence of absence of a Masters' degree explains approximately zero percent of variance in teacher quality, many states continue to require it if you want to keep your license, and almost every state will pay you more for having it.

Before taking my current job, I taught English in Japan. I had no Japanese language experience and no teaching experience, but the company I interviewed with asked if I had an undergraduate degree in some subject or other, and that was good enough for them. Meanwhile, I knew people who were fluent in Japanese and who had high-level TOEFL certification. They did not have a college degree so they were not considered.

My ex-girlfriend majored in Gender Studies, but it turned out all of the high-paying gender factories had relocated to China. They solved this problem by going to App Academy, a three month long, $15,000 course that taught programming. App Academy graduates compete for the same jobs as people who have taken computer science in college, a four year long, $200,000 undertaking.

I see no reason to think my family and friends are unique. The overall picture seems to be one of people paying hundreds of thousands of dollars to get a degree in Art History to pursue a job in Sales, or a degree in Spanish Literature to get a job as a middle manager. Or *not* paying hundreds of thousands of dollars, if they happen to be poor, and so being permanently locked out of jobs as a firefighter or salesman.

IV.

So presidential candidate Bernie Sanders has proposed universal free college tuition.

On the one hand, I sympathize with his goals. If you can't get any job better than 'fast food worker' without a college degree, and poor people can't afford college degrees, that's a pretty grim situation, and obviously unfair to the poor.

On the other hand, if can't you get married without a tulip, and poor people can't afford tulips, that's also a pretty grim situation, and obviously unfair to the poor.

But the solution isn't universal tulip subsidies.

Higher education is in a bubble much like the old tulip bubble. In the past forty years, the price of college has dectupled (quadrupled when adjusting for inflation). It used to be easy to pay for college with a summer job; now it is impossible. At the same time, the unemployment rate of people without college degrees is twice that of people who have them. Things are clearly very bad and Senator Sanders is right to be concerned.

But, well, when we require doctors to get a college degree before they can go to medical school, we're throwing out a mere $5 billion, barely enough to house all the homeless people in the country. But Senator Sanders admits that his plan would cost $70 billion per year. That's about the size of the

entire economy of Hawaii. It's enough to give $2000 every year to every American in poverty.

At what point do we say "Actually, no, let's not do that, and just let people hold basic jobs even if they don't cough up a a hundred thousand dollars from somewhere to get a degree in Medieval History"?

I'm afraid that Sanders' plan is a lot like the tulip subsidy idea that started off this post. It would subsidize the continuation of a useless tradition that has turned into a speculation bubble, prevent the bubble from ever popping, and disincentivize people from figuring out a way to route around the problem, eg replacing the tulips with daffodils.

(yes, it is nice to have college for non-economic reasons too, but let's be honest – if there were no such institution as college, would you, totally for non-economic reasons, suggest the government pay poor people $100,000 to get a degree in Medieval History? Also, anything not related to job-getting can be done three times as quickly by just reading a book.)

If I were Sanders, I'd propose a different strategy. Make "college degree" a protected characteristic, like race and religion and sexuality. If you're not allowed to ask a job candidate whether they're gay, you're not allowed to ask them whether they're a college graduate or not. You can give them all sorts of examinations, you can ask them their high school grades and SAT scores, you can ask their work history, but if you ask them if they have a degree then that's illegal class-based discrimination and you're going to jail. I realize this is a blatant violation of my usual semi-libertarian principles, but at this point I don't care.

SlateStarCodex Gives a Graduation Speech

[Trigger warning for deliberately provoking horror about graduates' real-world post-college prospects]
[Epistemic status: intended as persuasive speech, may somewhat overstate case]

Ladies and gentlemen, I am honored to have been invited to speak here at the great University of [mumble]. Go Wildcats, Spartans, or Eagles, as the case may be!

I apologize if what I have to say to you sounds a little unpolished. I was called in on very short notice after your original choice for graduation speaker, Mr. Steven L. Carter, had his invitation to speak rescinded due to his offensive and quite honestly outrageous opinions. Let me say in no uncertain terms that I totally condemn him and everything he stands for, and that I am glad to see the University of [mumble] taking a strong stand against this sort of thing.

Ladies and gentlemen, probably the most famous graduation speech in history was Kurt Vonnegut's "Wear Sunscreen" address. I'm sure you've all heard about it. He told an MIT class that they should wear sunscreen. Because for all he knew any more substantial advice he gave might be wrong, but that at least was on a firm evidential basis.

Well, I come here before you to explain that there is now serious controversy in the dermatological community. A 1995 paper found that people who used more sunscreen had a much higher risk of malignant melanoma, the most dangerous type of skin cancer. Eight years later, a review article claimed that the original paper was confounded by fairness of skin, and that likely the relationship between sunscreen use and melanoma is zero. But the story was further complicated by the finding that sunscreen use may increase cancers of the internal organs,

either through vitamin D dependent or some vitamin D independent pathways. My understanding is that a majority of dermatologists are still in favor of sunscreen, but that the issue is by no means settled.

But think about what the disagreement means. One of the smartest men in America came before an auditorium just like this, and said that there was only one item of advice of which he was completely certain – that you should wear sunscreen. Absolutely certain. And years later, we know that not only is this a very complicated question on which no certainty is yet possible – but it may very well be that if you follow his advice, you will get cancer and die.

Sometimes the things everybody knows everybody knows just aren't true. Like, did you know Vonnegut never wrote a graduation speech about sunscreen at all?

So with this spirit of questioning assumptions in mind, I want to ask you a question. Today many of you will be completing your education. Sure, some of you are going on to graduate or professional training, but it is clearly the end of an era. Seventeen years, from kindergarten to the present, and I want to ask you:

Is education worth it?

This sounds like the introduction to every college graduation speech ever. The speaker will ask if education is worth it, say of course it is because something something the human condition, and everyone will cheer and head off to the reception. So in order to keep you on your toes, I want to make the opposite point. What if education, as you understand it – public or private or charter schooling from age four or five all the way to university as young adults – is, on net, a waste of your time and money?

In order to move beyond platitudes in evaluate whether education is worthwhile – to give it the same kind of fair hearing we would want to give sunscreen – we need to list out some of the costs and benefits. Of benefits, two stand out clearly. The philosophical benefits of feeling connected to the beauty of mathematics, the passion of the humanities, the great historical traditions. And the practical benefits of being able to get a job and afford nice things like food and shelter.

We will start with philosophy. Human knowledge is pretty great. Your life has been enriched with the ideas of brilliant thinkers, of giants upon whose shoulders you might one day hope to stand. Isn't this enough?

But as 86% of you know, you can't just observe an experimental group has experienced an effect and attribute it to the experimental intervention. You have to see if other people in a control group got the same benefit for less work.

What would be the control group for school? Home-schoolers do much better than those who attend public or private schools by nearly any measure. But this is unfair; it's what scientists call an "active control". What we really need to do is compare you to people who got no instruction at all.

It's illegal not to educate a child, so our control group will be hard to find. But perhaps the best bet will be the "unschooling" movement, a group of parents who think school is oppressive and damaging. They *tell* the government they're home-schooling their children but actually just let them do whatever they want. They may teach their kid something if the child wants to be taught, otherwise they will leave them pretty much alone.

And this is really hard to study, because they're a highly self-selected group and there aren't very many of them. The only

study I could find on the movement only had n = 12, and although it tried as hard as it could to compare them to schoolchildren matched for race and family income level and parent education and all that good stuff I'm sure there's some weirdness that slipped through the cracks. Still, it's all we've got.

So, do these children do worse than their peers at public school?

Yes, they do.

By one grade level.

About college we still know very little. But if you'd stayed out of public school and stayed home and played games and maybe asked your parents some questions, then by the time your friends were graduating twelfth grade, you would have the equivalent of an eleventh-grade education.

Another intriguing clue here is Louis Benezet's experiment with mathematics instruction. Benezet, an early 20th century superintendent of schools, wondered whether cramming mathematics into kids at an early age had a detrimental effect. He decreed that in some of the schools in his district, there would be *no* math instruction until grade six. He found that within a year, these sixth graders had caught up with their peers in traditional schools, and furthermore that they were able to think much more logically about math problems – figure out what was going on rather than desperately trying to multiply and divide all the numbers in the problem by one another. If Benezet's results hold true – and on careful reading they are hard to doubt – any math education before grade six is useless *at best*. And it's hard to resist the urge to generalize to other subjects and children even older still.

Why is it so easy for the unschooled to keep up with their better educated brethren? My guess is that it's because very little learning goes on at school at all. The proponents of education speak of feeling connected to the beauty of mathematics, the passion of the humanities, and the great historical traditions. But how many of the children they spit out can prove one of Euclid's theorems? How many have been exposed to the Canterbury Tales? How many have experienced the sublime beauty of the Parthenon?

These aren't rhetorical questions, by the way. According to the general survey of knowledge among college students, 3.3% know who Euclid was, 7.6% know who wrote *Canterbury*, and a full 15% know what city the Parthenon's in.

36% of high school students know that an atom is bigger than an electron, rather than vice versa. But a full 59% of college students know the same. That's a whole nine percent better than chance. On one of the most basic facts about the fundamental entities that make up everything in existence.

"But knowledge isn't about names and dates!" No, but names and dates are the parts that are easy to measure, and it's a pretty good bet that if you don't know what city the Parthenon's in you probably haven't absorbed the full genius of the Greek architectural tradition. Anyone who's never heard of Chaucer probably doesn't have strong opinions on the classics of Middle English literature.

So in contradiction to the claim that education is necessary to teach beautiful and elegant knowledge, I maintain first that nearly nobody in the educational system picks this up anyway, that people who don't get any formal education at all pick it up nearly as much of it, and that people not exposed to it as children will, if they decide to learn it as adults, pick it up

quickly and easily and without the heartbreak of trying to cram it into the underdeveloped head of a seven year old.

What about the claim that education is practically useful for getting a job and making money?

Even more than most young people, you've had the privilege of getting to watch your dreams implode in real time right before your eyes. About fifteen percent of you will be some variant of unemployed straight out of college. Another ten percent will find something part-time. And another forty or so percent will be underemployed, working as waiters or clerks or baristas or something else that uses zero percent of the knowledge you've worked so hard to accumulate. The remaining third of you who get something vaguely resembling the job you signed up for will still have to deal with wages that have stagnated over the last decade even as working hours increased and average student debt nearly doubled.

But don't worry, I'm sure the nice folks at Chase-Bear-Goldman-Sallie-Manhattan-Stearns-Sachs-Mae-FEDGOV will be happy to forgive your debt if you mention you weren't entirely happy with the purchase. You did hold out for the satisfaction-guaranteed offer, right? No? Uh oh.

As bad as the job market is, staying in school looks worse. Economists warn that attending law school is the worst career decision you can make, so much so that newly graduated lawyers have nothing do to but sue law schools for not warning them against attending and established firms offer an Anything But Law School Scholarship to raise awareness of the problem. Doctors are so uniformly unhappy that they are committing suicide in record numbers and nine out of ten would warn young people against going into medicine. Graduate school has always been an iffy bet, but now the ratio

of Ph. D applicants to open tenure track positions has hit triple digits, with the vast majority ending up as miserable adjunct professors who juggle multiple part time jobs and end up making as much as a Starbucks barista but without the health insurance.

I'd like to thank whoever figured out how to include URLs in speeches, by the way. That was *the best* invention.

But here I cannot honestly disagree with the conventional assessment that going to school raises your earning power. As bad as you will have it, everyone who didn't graduate college still has it much, much worse. All the economic indicators agree with the signs from the desolate wasteland that was once our industrial heartland: they are doomed. Their wages are not stagnating but actively declining, their unemployment rate is a positively Greek thirty-five percent, and prospects for changing that are few and far between. Some economists blame globalization, which makes it easy to outsource manufacturing and other manual labor to the Chinese. Others blame technology, noting that many of the old well-paying blue-collar jobs are done not by foreigners but by machines. Both trends are set to increase, turning even more factory workers, truck drivers, and warehouse-stockers into burger-flippers, Wal-Mart greeters, and hollow-eyed unemployed.

But don't let your schadenfreude get the better of you. Twenty years from now that's going to be you. Sure, right now machines can only do the easy stuff, and the world isn't interconnected enough to let foreigners do anything really *subtle* for us. But lawyers are already feeling the pinch of software that auto-generates contracts, and programmers are already feeling the pinch of Indians who will work for half the pay and email their code to Silicon Valley the next morning. You don't need to invent a robo-drafter to put engineers out of

business, just drafting software so effective it allows one engineer to do the work of three. And although there are half-hearted efforts to stop it, it seems more and more like King Canute trying to turn back a tide made of hundred dollar bills.

Once machines can do everything we can better and cheaper, the inevitable end result is employment for a few geniuses who invent and run the machines, immense profits for the capitalists who own the machines, and what happens to everyone else better left unspoken.

"Is this a vision of what shall be, or of what might be only?" Well, a visionaries as diverse as Martin Luther King, Richard Nixon and Milton Friedman have proposed something called a Basic Income Guarantee. When society becomes so advanced that it produces more than enough for everybody – but also so advanced that most individuals below genius level have little to contribute and no way of earning money – everyone should get a yearly salary just for existing. Think welfare, except that it goes to everybody, there's no stigma, and it's more than enough to live on. This titanic promise has run up against a giant iceberg with BUT HOW WOULD WE PAY FOR IT written in big red letters on the front. If we cancelled all existing welfare and entitlement programs – which makes sense if we're giving everyone enough money to live comfortably on, we would only free up enough money together for a universal income of $5,800. I don't know if you can live on that, but I'd hate to have to try.

But we've gotten off track. We were counting the benefits of formal education. We did not do so well in trying to prove that it left you more knowledgeable, but it did seem like it had some practical value in getting you a little bit more money. With your shiny college degree, you can confidently assert "I've got mine", just as long as you take care not to notice the

increasingly distant hordes of manual laborers or the statistics showing that the yours you've got is less and less every year.

What of the costs of education? What have you lost out on?

Well, first about twenty thousand hours of your youth. That's okay. You weren't using that golden time of perfect health and halcyon memories when you had more true capacity for creativity and imagination and happiness than you ever will again anyway. If you hadn't had your teachers to tell you that you needed to be making a collage showing your feelings about *The Scarlet Letter*, you probably would have wasted your childhood seeing a world in a grain of sand or Heaven in a wild flower or something dumb like that.

I'm more interested in the financial side of it. At $11,000 average per pupil spending per year times thirteen years plus various preschool and college subsidies, the government spends $155,000 on the kindergarten-through-college education of the average American.

Inspired by a tweet: what if the government had taken this figure (adjusted for inflation) and invested it in the stock market at the moment of your birth? Today when you graduate college, they remove it from the stock market, put it in a low-risk bond, put a certain percent of the interest from that bond into keeping up with inflation, and hand you the rest each year as a basic income guarantee. How much would you have?

And I calculate that the answer would be $15,000 a year, adjusted for interest. We can add the $5,800 basic income guarantee we could already afford onto that for about $20,000 a year, for everyone. Black, white, man, woman, employed, unemployed, abled, disabled, rich, poor. Welcome to the real world, it's dangerous to go alone, take this. What, you thought we were going to throw you out to sink or swim in a world

where if you die *you die in real life?* Come on, we're not that cruel.

So when we ask whether your education is worth it, we have to compare what you got – an education that puts you one grade level above the uneducated and which has informed 3.3% of you who Euclid is – to what you could have gotten. 20,000 hours of your youth to play, study, learn to play the violin, whatever. And $20,000 a year, sweat-free.

$20,000 a year isn't much. The average mid-career salary of an average college graduate is nearly triple that – $55,000. By the numbers your education looks pretty good. But numbers can be deceiving.

Consider the life you have to look forward to, making your $55,000. The exact profession that makes closest to that number is a paralegal, so let's go with that. You get a job as a paralegal in a prestigious Manhattan law firm. You can't afford to live in Manhattan, but you scrounge together enough money for a cramped apartment in Brooklyn, which costs you about $2000 a month rent. Every morning you wake up at 7:45, get on the forty-five minute subway ride to Manhattan, and make it to work by your 9:00 AM starting time. Your boss is a kind of nasty lawyer who is himself upset that he can't pay back his law school debt and yells at you all day. By the time you get back home around 6, you're too exhausted to do much besides watch some TV. You don't really have time to meet guys – I'm assuming you're a woman here, sixty percent of you are, I blame the patriarchy – so you put out a personal ad on Craigslist and after a while find someone you like. You get married after a year; your honeymoon is in Vermont because his company won't give him enough time off to go any further.

You have two point four kids, and realize you've got to move to a better part of town because your school district sucks. Combined with your student debt, that puts a big strain on the finances and you don't have enough to pay for child care. Eventually you find a place that will do it for cheap, and although it looks kind of dirty and you're shocked when Junior calls you a "puta" which isn't even a proper *English* curse word the price is right and they're the only people who will accept four tenths of a kid. The older kids keep asking you and Dad for help with homework, which you can't give because you haven't really had time to keep up with your math and grammar and so on skills, what with the paralegal job and the television-watching taking up all your time. So you tell them to ask their teacher for extra help, which their teacher doesn't give because she's got forty other kids asking for the same thing and only twenty-four hours in a day. Despite all of this Junior gets into college and *you* sure haven't saved up the money to put him through there tuition has spiraled to twelve gazillion dollars by this point and Chase-Bear-Goldman-Sallie-Manhattan-Stearns-Sachs-Mae-FEDGOV can't lend him that because gazillion isn't even a real number, and ohmigod what if Junior ends up one of those high school graduates with the Greek-level unemployment rates standing forlornly in front of a decaying factory in the Rust Belt? Worse, what if he ends up *living with you*? You beg him to go back to the bank and offer to pay whatever interest rates they ask. And so the cycle begins anew.

Or consider your life on a $20,000 a year income guarantee. No longer tied down to a job, you can live wherever you want. I love the mountains. Let's live in a cabin in Colorado, way up in the Rockies. You can find stunningly beautiful ones for $500 a month – freed from the mad rush to get into scarce

urban or suburban areas with good school districts, housing is actually really cheap. So there you are in the Rockies, maybe with a used car to take you to Denver when you want to see people or go to a show, but otherwise all on your own except for the deer and squirrels. You wake up at nine, cook yourself a healthy breakfast, then take a long jog out in the forest. By the time you come back, you've got a lot of interesting thoughts, and you talk about them with the dozens of online friends you cultivate close relationships with and whom you can take a road trip and visit any time you feel like. Eventually you're talked out, and you curl up with a good book – this week you're trying to make it through Aristotle on aesthetics. The topic interests you since you're learning to paint – you've always wanted to be an artist, and with all the time in the world and stunning views to inspire you, you're making good progress. Freed from the need to appeal to customers or critics, you are able to develop your own original style, and you take heart in the words of the old Kipling poem:

> And none but the Master will praise them
> And none but the Master will blame
> And no one will work for money
> And no one will work for fame
> But each for the joy of the working
> Each on his separate star
> To draw the thing as he sees it
> For the God of things as they are

One of the fans of your work is a cute girl – this time I'm assuming you're a man, I'm sure over the past four years you've learned some choice words for people who do that. You date and get married. She comes to live with you – she's also getting $20,000 a year from the government in place of an

education, so now you're up to $40,000, which is actually very close to the US median household income. You have two point four kids. With both of you at home full time, you see their first steps, hear their first words, get to see them as they begin to develop their own personalities. They start seeming a little lonely for other kids their own age, so with a sad good-bye to your mountain, you move to a bigger house in a little town on the shores of a lake in Montana. There's no schooling for them, but you teach them to read, first out of children's books, later out of something a little harder like Harry Potter, and then finally you turn them loose in your library. Your oldest devours your collection of Aristotle and tells you she wants to be a philosopher when she grows up. Evenings they go swimming, or play stickball with the other kids in town.

When they reach college age, your daughter is so thrilled at the opportunity to learn from her intellectual heroes that she goes to Chase-Bear-Goldman-Sallie-Manhattan-Stearns-Sachs-Mae-FEDGOV and asks for a loan. They're happy to give her fifteen thousand, which is all college costs nowadays – only the people who are really interested in learning feel the need to go nowadays, and supply so outpaces demand that prices are driven down. She makes it into Yale (unsurprising given how much better home-schooled students do) studies philosophy, but finds she likes technology better. She decides to become an engineer, and becomes part of the base of wealthy professionals helping fund the income guarantee for everyone else. She marries a nice man *after* making sure he's willing to stay home and take care of the children – she's not crazy, she doesn't want to send them to some kind of *institution*

Your younger son, on the other hand, is a little intellectually disabled and can't read above a third-grade level. That's not a

big problem for you or for him. When he grows older, he moves to Hawaii where he spends most of his time swimming in the ocean and by all accounts enjoys himself very much.

You're happy your son will be financially secure for the rest of his life, but on a broader scale, you're happy that no one around you has to live in fear of getting fired, or is struggling to make ends meet, or is stuck in the Rust Belt with a useless skill set. Every so often, you call your daughter and thank her for helping design the robots that do most of the hard work.

Would you like to swing on a star? Carry moonbeams home in a jar? And be better off than you are? Or would you like to get a formal education?

We're finally getting back to the point now. I'm sorry it's taken this long. I can see the Dean of Students checking her watch over there with a worried look on her face. I think she's worried I'm trying to filibuster your graduation. You know legally if I can keep speaking until midnight tonight, the graduation is cancelled and you have to stay in school another year? It's true. Those are the rules.

Because I don't want to talk about the very broad social question of whether Education the concept is worth it to Society as a concept. I want to ask *you*, standing here today, was *your* education worth it?

Because this is a college graduation speech, and I am legally mandated to offer some advice, and the specific advice I give will be tailored to your response.

Some of you will say yes, my education was worth it. I am the 3.3%! I know who Euclid was and I understand the sublime beauty of geometry. I don't think I would have been exposed to it, or had the grit to keep studying it, if I hadn't been here surrounded by equally curious peers, under the instruction of

enthusiastic professors. This revelation was worth losing my cabin in Colorado, worth resigning myself to the daily grind and the constant lurking fear of failure. I claim it all.

And to you my advice is: if you've sacrificed everything for knowledge, don't forget that. When you are a paralegal in Brooklyn, and you get home from work, and you are very tired, and you want to curl up in front of the TV and watch reality shows until you are numb, remind yourself that you value knowledge above everything else, that you will seek intellectual beauty though the world perish, and read a book or something. Or take a class at a community college. Anything other than declaring knowledge your supreme value but becoming a boob.

Others of you will say yes, my education was worth it. Not because of what I learned about ukulele or eucalyptus or whatever, but because of the friends I made here, the proud University of [mumble] spirit of camaraderie, which I will carry forth my entire life.

And to you my advice is similar: if you've sacrificed everything for friendship, don't forget that. When you are a paralegal in Brooklyn, or a market analyst in Seattle, or God forbid an intern in Michigan, and you get home from work, and you are very tired, and you want to curl up in front of your computer and check Reddit, remind yourself of the friends you made here and give them a call. See how they're doing. Write them a Christmas card, especially if it is December. Anything other than declaring friendship your supreme value and drifting out of touch.

Others of you will say yes, my education was worth it. Not because of what I learned about the Eucharist or eucre or whatever, but because of the connections I made, the network

of alumni who will be giving me a leg up in whatever I choose to pursue.

And to you my advice is, again, similar. If you've sacrificed everything for ambition, be ambitious as *hell*. When you are a paralegal in Brooklyn or whatever, claw your way to the top, stay there, and use it to do something important. If you've sacrificed everything for ambition, don't you dare stop at middle manager.

Others of you will say yes, my education was worth it. Not because of what I learned about yucca or the Yucatan or whatever, but because it helped me learn civic values, become a better person who is better able to help others.

And to you my advice is once again similar. If you've sacrificed everything to help others, don't let it all end with donating a tenner to the OXFAM guy on the street now and then. Join Giving What We Can or go volunteer somewhere. If you've sacrificed everything for others, make sure others get something good out of the deal!

Others of you will say yes, my education was worth it. Not because of what I learned about eukaryotes or Ukraine or whatever, but because formal education in the school system *taught me how to think*.

And to…sorry, one second, HAHAHAHAHAHHAHAAAHAHAHAHHHAAHAHA HAHAHAHAHAHHA HAHAHAHAHAHAH HAHAHAHHHHHAAAHAHAHAHAH HAHAHHA HAHAHHHHAHAH HAAHHHAHA HAAHAHAHAHHA HAHHAHAHAHAHAHHAHA AHHHHAHAHAHAHA HAHAHAHAHA HHAAAHAHHAAHHAHA AHHAHAHAHAHA hahaha haha ha hahaha haha heh heh heh okay.

I'm sorry. Ahem. To you my advice is, again, similar. If you've sacrificed everything to learn how to think, learn how to think. When someone says something you disagree with, before you dismiss a straw man it and call that person names and slap yourself five for your brilliant rebuttal, take a second to consider it fairly on its own terms. Go learn about biases and heuristics and how to avoid them. Read enough psychology and cognitive science to figure out why your claim might *kind of* inspire hysterical laughter from people even a little familiar with the field. Just don't sacrifice everything to learn how to think and end up only rearranging your prejudices.

And finally, some of you will say, wait a second, maybe my education *wasn't* worth it. Or, maybe it was the best choice to make from within a bad paradigm, but I'm not content with that. And I wish someone had told me about all of this more than fifteen minutes before I graduate.

And to you I can offer a small amount of compensation. You have learned a very valuable lesson that you might not have been able to learn any other way.

You have learned that the system is Not Your Friend.

I use those last three words very consciously. People usually say "not your friend" as an understatement, a way of saying something is actively hostile. I don't mean that.

The system is not your friend. The system is not your enemy. The system is a retarded giant throwing wads of $100 bills and books of rules in random directions while shouting "LOOK AT ME! I'M HELPING! I'M HELPING!" Sometimes by luck you catch a wad of cash, and you think the system loves you. Other times by misfortune you get hit in the gut with a

rulebook, and you think the system hates you. But either one is giving the system too much credit.

Every one of the architects and leaders of the system is fantastically intelligent – some even have degrees from the University of [mumble]. But every one of the neurons in my dog's brain is a fantastically complex pinnacle of three billion years of evolution, yet my dog herself can spend the better part of an hour standing motionless, hackles raised, barking at a plastic bag.

To you I don't have very much advice. I'm no smarter than anyone else – well, I know who Euclid is, but *other* than that – and if I knew how to fix the system, it's a pretty good bet other people would know too and the system would already have been fixed. Maybe you, armed with a degree from the University of [mumble], will be the one to help figure it out.

On the other hand, someone a lot smarter than I am *did* have some advice for you. Poor Kurt Vonnegut never did get to give a real graduation speech, but one of his books has some advice targeted at another major life transition:

> Hello babies. Welcome to Earth. It's hot in the summer and cold in the winter. It's round and wet and crowded. On the outside, babies, you've got a hundred years here. There's only one rule that I know of, babies-"God damn it, you've got to be kind."

I don't know how to fix the system, but I am pretty sure that one of the ingredients is kindness.

I think of kindness not only as the moral virtue of volunteering at a soup kitchen or even of living your life to help as many other people as possible, but also as an epistemic virtue. Epistemic kindness is kind of like humility. Kindness to ideas

you disagree with. Kindness to positions you want to dismiss as crazy and dismiss with insults and mockery. Kindness that breaks you out of your own arrogance, makes you realize the truth is more important than your own glorification, especially when there's a lot at stake.

Here we are at the end of a grinder of $150,000, 20,000 hours, however many dozen collages about *The Scarlet Letter*, and the occasional locker room cry of "faggot" followed by a punch in the gut. Somewhere in another world, there are people just like us in nice cabins reading Aristotle and knowing that nobody will have to go hungry ever again. The difference between us and them isn't money, because I think the $155,000 the government gave you could have gone either way – and even if I'm wrong about that there's more than enough money somewhere else. The difference isn't intelligence, because the architects of our system are fantastically bright in their own way. I think kindness might be that difference.

Technically kindness plus coordination power, but that's another speech, and the Dean of Students is starting to make frantic hand signals.

I don't know if it's really possible to afford to give everyone that cabin in Colorado. But I hope that the people whose job it is to figure that out approach the problem with a spirit of kindness and humility.

In conclusion, both sides of the sunscreen debate have some pretty good points. It will certainly decrease your risk of squamous and basal cell carcinomas, it probably has no effect on the malignant melanoma rate but there's a nonzero chance it might either cause *or* prevent them, and its effect on internal

tumors seems worrying at this point but is yet to be backed up by any really firm evidence.

I understand this is complicated and unsatisfying. Welcome to the real world.

[Congratulations to my girlfriend Ozy, who graduates college this week!]

X. Progress

Intellectual Hipsters and Meta-Contrarianism

Related to: Why Real Men Wear Pink, That Other Kind of Status, Pretending to be Wise, The "Outside The Box" Box

> *WARNING: Beware of things that are fun to argue — Eliezer Yudkowsky*

Science has inexplicably failed to come up with a precise definition of "hipster", but from my limited understanding a hipster is a person who deliberately uses unpopular, obsolete, or obscure styles and preferences in an attempt to be "cooler" than the mainstream. But why would being deliberately uncool be cooler than being cool?

As previously discussed, in certain situations refusing to signal can be a sign of high status. Thorstein Veblen invented the term "conspicuous consumption" to refer to the showy spending habits of the nouveau riche, who unlike the established money of his day took great pains to signal their wealth by buying fast cars, expensive clothes, and shiny jewelery. Why was such flashiness common among new money but not old? Because the old money was so secure in their position that it never even occurred to them that they might be confused with poor people, whereas new money, with their lack of aristocratic breeding, worried they might be mistaken for poor people if they didn't make it blatantly obvious that they had expensive things.

The old money might have started off not buying flashy things for pragmatic reasons - they didn't need to, so why waste the money? But if F. Scott Fitzgerald is to be believed, the old money actively cultivated an air of superiority to the nouveau

riche and their conspicuous consumption; not buying flashy objects becomes a matter of principle. This makes sense: the nouveau riche need to differentiate themselves from the poor, but the old money need to differentiate themselves from the nouveau riche.

This process is called countersignaling, and one can find its telltale patterns in many walks of life. Those who study human romantic attraction warn men not to "come on too strong", and this has similarities to the nouveau riche example. A total loser might come up to a woman without a hint of romance, promise her nothing, and demand sex. A more sophisticated man might buy roses for a woman, write her love poetry, hover on her every wish, et cetera; this signifies that he is not a total loser. But the most desirable men may deliberately avoid doing nice things for women in an attempt to signal they are so high status that they don't need to. The average man tries to differentiate himself from the total loser by being nice; the extremely attractive man tries to differentiate himself from the average man by not being especially nice.

In all three examples, people at the top of the pyramid end up displaying characteristics similar to those at the bottom. Hipsters deliberately wear the same clothes uncool people wear. Families with old money don't wear much more jewelry than the middle class. And very attractive men approach women with the same lack of subtlety a total loser would use.[1]

If politics, philosophy, and religion are really about signaling, we should expect to find countersignaling there as well.

Pretending To Be Wise

Let's go back to Less Wrong's long-running discussion on death. Ask any five year old child, and ey can tell you that

death is bad. Death is bad because it kills you. There is nothing subtle about it, and there does not need to be. Death universally seems bad to pretty much everyone on first analysis, and what it seems, it is.

But as has been pointed out, along with the gigantic cost, death does have a few small benefits. It lowers overpopulation, it allows the new generation to develop free from interference by their elders, it provides motivation to get things done quickly. Precisely because these benefits are so much smaller than the cost, they are hard to notice. It takes a particularly subtle and clever mind to think them up. Any idiot can tell you why death is bad, but it takes a very particular sort of idiot to believe that death might be good.

So pointing out this contrarian position, that death has some benefits, is potentially a signal of high intelligence. It is not a very reliable signal, because once the first person brings it up everyone can just copy it, but it is a cheap signal. And to the sort of person who might not be clever enough to come up with the benefits of death themselves, and only notices that wise people seem to mention death can have benefits, it might seem super extra wise to say death has lots and lots of great benefits, and is really quite a good thing, and if other people should protest that death is bad, well, that's an opinion a five year old child could come up with, and so clearly that person is no smarter than a five year old child. Thus Eliezer's title for this mentality, "Pretending To Be Wise".

If dwelling on the benefits of a great evil is not your thing, you can also pretend to be wise by dwelling on the costs of a great good. All things considered, modern industrial civilization - with its advanced technology, its high standard of living, and its lack of typhoid fever - is pretty neat. But modern industrial civilization also has many costs: alienation from nature, strains

on the traditional family, the anonymity of big city life, pollution and overcrowding. These are real costs, and they are certainly worth taking seriously; nevertheless, the crowds of emigrants trying to get from the Third World to the First, and the lack of any crowd in the opposite direction, suggest the benefits outweigh the costs. But in my estimation - and speak up if you disagree - people spend a lot more time dwelling on the negatives than on the positives, and most people I meet coming back from a Third World country have to talk about how much more authentic their way of life is and how much we could learn from them. This sort of talk sounds Wise, whereas talk about how nice it is to have buses that don't break down every half mile sounds trivial and selfish..

So my hypothesis is that if a certain side of an issue has very obvious points in support of it, and the other side of an issue relies on much more subtle points that the average person might not be expected to grasp, then adopting the second side of the issue will become a signal for intelligence, even if that side of the argument is wrong.

This only works in issues which are so muddled to begin with that there is no fact of the matter, or where the fact of the matter is difficult to tease out: so no one tries to signal intelligence by saying that 1+1 equals 3 (although it would not surprise me to find a philosopher who says truth is relative and this equation is a legitimate form of discourse).

Meta-Contrarians Are Intellectual Hipsters

A person who is somewhat upper-class will conspicuously signal eir wealth by buying difficult-to-obtain goods. A person who is very upper-class will conspicuously signal that ey feels no need to conspicuously signal eir wealth, by deliberately not buying difficult-to-obtain goods.

A person who is somewhat intelligent will conspicuously signal eir intelligence by holding difficult-to-understand opinions. A person who is very intelligent will conspicuously signal that ey feels no need to conspicuously signal eir intelligence, by deliberately not holding difficult-to-understand opinions.

According to the survey, the average IQ on this site is around 145[2]. People on this site differ from the mainstream in that they are more willing to say death is bad, more willing to say that science, capitalism, and the like are good, and less willing to say that there's some deep philosophical sense in which 1+1 = 3. That suggests people around that level of intelligence have reached the point where they no longer feel it necessary to differentiate themselves from the sort of people who aren't smart enough to understand that there might be side benefits to death. Instead, they are at the level where they want to differentiate themselves from the somewhat smarter people who think the side benefits to death are great. They are, basically, meta-contrarians, who counter-signal by holding opinions contrary to those of the contrarians' signals. And in the case of death, this cannot but be a good thing.

But just as contrarians risk becoming too contrary, moving from "actually, death has a few side benefits" to "DEATH IS GREAT!", meta-contrarians are at risk of becoming too meta-contrary.

All the possible examples here are controversial, so I will just take the least controversial one I can think of and beg forgiveness. A naive person might think that industrial production is an absolute good thing. Someone smarter than that naive person might realize that global warming is a strong negative to industrial production and desperately needs to be stopped. Someone even smarter than that, to differentiate

emself from the second person, might decide global warming wasn't such a big deal after all, or doesn't exist, or isn't man-made.

In this case, the contrarian position happened to be right (well, maybe), and the third person's meta-contrariness took em further from the truth. I do feel like there are more global warming skeptics among what Eliezer called "the atheist/libertarian/technophile/sf-fan/early-adopter/programmer empirical cluster in personspace" than among, say, college professors.

In fact, very often, the uneducated position of the five year old child may be deeply flawed and the contrarian position a necessary correction to those flaws. This makes meta-contrarianism a very dangerous business.

Remember, most everyone hates hipsters.

Without meaning to imply anything about whether or not any of these positions are correct or not[3], the following triads come to mind as connected to an uneducated/contrarian/meta-contrarian divide:

- KKK-style racist / politically correct liberal / "but there are scientifically proven genetic differences"
- misogyny / women's rights movement / men's rights movement
- conservative / liberal / libertarian[4]
- herbal-spiritual-alternative medicine / conventional medicine / Robin Hanson
- don't care about Africa / give aid to Africa / don't give aid to Africa
- Obama is Muslim / Obama is obviously not Muslim, you idiot / Patri Friedman[5]

What is interesting about these triads is not that people hold the positions (which could be expected by chance) but that people get deep personal satisfaction from arguing the positions even when their arguments are unlikely to change policy[6] - and that people identify with these positions to the point where arguments about them can become personal.

If meta-contrarianism is a real tendency in over-intelligent people, it doesn't mean they should immediately abandon their beliefs; that would just be meta-meta-contrarianism. It means that they need to recognize the meta-contrarian tendency within themselves and so be extra suspicious and careful about a desire to believe something contrary to the prevailing contrarian wisdom, especially if they really enjoy doing so.

Footnotes

1) But what's really interesting here is that people at each level of the pyramid don't just follow the customs of their level. They enjoy following the customs, it makes them feel good to talk about how they follow the customs, and they devote quite a bit of energy to insulting the people on the other levels. For example, old money call the nouveau riche "crass", and men who don't need to pursue women call those who do "chumps". Whenever holding a position makes you feel superior and is fun to talk about, that's a good sign that the position is not just practical, but signaling related.

2) There is no need to point out just how unlikely it is that such a number is correct, nor how unscientific the survey was.

3) One more time: *the fact that those beliefs are in an order does not mean some of them are good and others are bad*. For example, "5 year old child / pro-death / transhumanist" is a triad, and "warming denier / warming believer / warming

skeptic" is a triad, but I personally support 1+3 in the first triad and 2 in the second. You can't evaluate the truth of a statement by its position in a signaling game; otherwise you could use human psychology to figure out if global warming is real!

4) This is my solution to the eternal question of why libertarians are always more hostile toward liberals, even though they have just about as many points of real disagreement with the conservatives.

5) To be fair to Patri, he admitted that those two posts were "trolling", but I think the fact that he derived so much enjoyment from trolling in that particular way is significant.

6) Worth a footnote: I think in a lot of issues, the original uneducated position has disappeared, or been relegated to a few rednecks in some remote corner of the world, and so meta-contrarians simply look like contrarians. I think it's important to keep the terminology, because most contrarians retain a psychology of feeling like they are being contrarian, even after they are the new norm. But my only evidence for this is introspection, so it might be false.

A Signaling Theory of Class x Politics Interaction

The media, most recently The Economist and Scientific American, have been publicizing a surprising statistical finding: in the current economic climate, when more Americans than ever are poor, support for policies that redistribute wealth to the poor are at their *lowest* levels ever. This new-found antipathy towards aid to the poor concentrates in people who are near but not yet on the lowest rung of the social ladder. The Economist adds some related statistics: those who earn slightly more than the minimum wage are most against raising the minimum wage, and support for welfare in an area decreases as the percentage of welfare recipients in the area rises.

Both articles explain the paradoxical findings by appealing to something called "last place aversion", an observed tendency for people to overvalue not being in last place. For example, in laboratory experiments where everyone gets randomly determined amounts of money, most people are willing to help those with less money than themselves gain cash - except the person with the second to lowest amount of money, who tends to try to thwart the person in last place even if it means enriching those who already have the most.

"Last place aversion" is interesting, and certainly deserves at least a footnote in the catalogue of cognitive biases and heuristics, but I find it an unsatisfying explanation for the observations about US attitudes toward wealth redistribution. For one thing, the entire point of last place aversion is that it only affects those in last place, but in a massive country like the United States, everyone can find someone worse off than

themselves (with one exception). For another, redistributive policies usually stop short of making those who need government handouts wealthier than those who do not; subsidizing more homeless shelters doesn't risk giving the homeless a nicer house than your own. Finally, many of the policies people oppose, like taxing the rich, don't directly translate to helping those in last place.

I propose a different mechanism, one based on … wait for it … signaling.

In a previous post, I discussed multi-level signaling and counter-signaling, where each level tries to differentiate itself from the level beneath it. For example, the nouveau riche differentiate themselves from the middle class by buying ostentatious bling, and the nobility (who are at no risk of being mistaken for the middle class) differentiate themselves from the nouveau riche by *not* buying ostentatious bling.

The very poor have one strong incentive to support redistribution of wealth: they need the money. They also have a second, subtler incentive: most redistributive policies come packaged with a philosophy that the poor are not personally responsible for the poverty, but are at least partially the victims of the rest of society. Therefore, these policies inflate both their pocketbook and their ego.

The lower middle class gain what status they have by not being the very poor; effective status signaling for a lower middle class person is that which proves that she is certainly not poor. One effective method is to hold opinions contrary to those of the poor: that redistribution of wealth is evil and that the poor deserve their poverty. This ideology celebrates the superiority of the lower middle class over the poor by emphasizing the biggest difference between the lower middle

class and the very poor: self-reliance. By asserting this ideology, a lower middle class person can prove her lower middle class status.

The upper middle class gain what status they have by not being the lower middle class; effective status signaling for an upper middle class person is that which proves that she is certainly not lower middle class. One effective way is to hold opinions contrary to those of the lower middle class: that really the poor and lower middle class are the same sort of people, but some of them got lucky and some of them got unlucky. The only people who can comfortably say "Deep down there's really no difference between myself and a poor person" are people confident that no one will *actually* mistake them for a poor person after they say this.

As a thought experiment, imagine your reactions to the following figures:

1. A bearded grizzled man in ripped jeans, smelling slightly of alcohol, ranting about how the government needs to give more free benefits to the poor.

2. A bearded grizzled man in ripped jeans, smelling slightly of alcohol, ranting about how the poor are lazy and he worked hard to get where he is today.

3. A well-dressed, stylish man in a business suit, ranting about how the government needs to give more free benefits to the poor.

4. A well-dressed, stylish man in a business suit, ranting about how the poor are lazy and he worked hard to get where he is today.

My gut reactions are (1, lazy guy who wants free money) (2, honorable working class salt-of-the-earth) (3, compassionate

guy with good intentions) (4, insensitive guy who doesn't realize his privilege). If these are relatively common reactions, these would suffice to explain the signaling patterns in these demographics.

If this were true, it would explain the unusual trends cited in the first paragraph. An area where welfare became more common would see support for welfare drop, as it became more and more necessary for people to signal that they themselves were not welfare recipients. Support for minimum wage would be lowest among people who earn just slightly more than minimum wage, and who need to signal that they are not minimum wage earners. And since upper middle class people tend to favor redistribution as a status signal and lower middle class people tend to oppose it, a recession that drives more people into the lower middle class would cause a drop in support for redistributive policies.

Reactionary Philosophy in an Enormous, Planet-Sized Nutshell

I have heard the following from a bunch of people, one of whom was me six months ago: “I keep on reading all these posts by really smart people who identify as Reactionaries, and I don’t have any idea what’s going on. They seem to be saying things that are either morally repugnant or utterly ridiculous. And when I ask them to explain, they say it’s complicated and there’s no one summary of their ideas. Why don’t they just write one?”

Part of me secretly thinks part of the answer is that a lot of these beliefs are not argument but poetry. Try to give a quick summary of Shelley’s *Adonais*: “Well there’s this guy, and he’s dead, and now this other guy is really sad.” One worries something has been lost. And just as well try to give a quick summary of the sweeping elegaic paeans to a bygone age of high culture and noble virtues that is Reaction.

But there *is* some content, and some of it is disconcerting. I started reading a little about Reaction after incessantly being sent links to various Mencius Moldbug posts, and then started hanging out in an IRC channel with a few Reactionaries (including the infamous Konkvistador) whom I could question about it. Obviously this makes me the world expert who is completely qualified to embark on the hitherto unattempted project of explaining it to everyone else.

Okay, maybe not. But the fact is, I’ve been itching to prsent an argument against Reactionary thought for a long time, but have been faced with the dual problem of not really having a solid target and worrying that everyone not a Reactionary would think I was wasting my time even talking to them.

Trying to sum up their ideas seems like a good way to first of all get a reference point for what their ideas are, and second of all to make it clearer why I think they deserve a rebuttal.

We'll start with the meta-level question of how confident we should be that our society is better than its predecessors in important ways. Then we'll look on the object level about how we compare to past societies along dimensions we might care about. We'll make a lengthy digression into social justice issues, showing how some traditional societies were actually more enlightened than our own in this area. Having judged past societies positively, we'll then look at what aspects of their cultures, governments, and religions made them so successful, and whether we could adopt those to modern life.

Much of this will be highly politically incorrect and offensive, because that's what Reactionaries *do*. I have tried to be charitable towards these ideas, which means this post will be pushing politically incorrect and offensive positions. If you do not want to read it, especially the middle parts which are about race, I would totally understand that. But if you do read it and accuse me of holding these ideas myself and get really angry, then you fail at reading comprehension forever.

I originally planned to follow this up tomorrow with the post containing my arguments against these positions, but this argument took longer than I thought to write and I expect the counterargument will as well. Expect a post critiquing reactionary ideas sometime in the next…week? month?

[EDIT: The Anti-Reactionary FAQ is now available]

In any case, this is not that post. This is the post where I argue that modern society is rotten to the core, and that the only reasonable solution is to dig up King James II, clone him, and give the clone absolute control over everything.

_ * _ * _ * _

No One Expects The Spanish Inquisition, Especially Not In 21st Century America

People in ancient societies thought their societies were obviously great. The imperial Chinese thought nothing could beat imperial China, the medieval Spaniards thought medieval Spain was a singularly impressive example of perfection, and Communist Soviets were pretty big on Soviet Communism. Meanwhile, we think 21st-century Western civilization, with its democracy, secularism, and ethnic tolerance is pretty neat. Since the first three examples now seem laughably wrong, we should be suspicious of the hypothesis that *we* finally live in the one era whose claim to have gotten political philosophy right is *totally justified*.

But it seems like we have an advantage they don't. Speak out against the Chinese Empire and you lose your head. Speak out against the King of Spain and you face the Inquisition. Speak out against Comrade Stalin and you get sent to Siberia. The great thing about western liberal democracy is that it has a free marketplace of ideas. *Everybody* criticizes some aspect of our society. Noam Chomsky made a career of criticizing our society and became rich and famous and got a cushy professorship. So our advantage is that we admit our society's imperfections, reward those who point them out, and so keep inching closer and closer to this ideal of perfect government.

Okay, back up. Suppose you went back to Stalinist Russia and you said "You know, people just don't respect Comrade Stalin enough. There isn't enough Stalinism in this country! I say we need *two* Stalins! No, *fifty* Stalins!"

Congratulations. You have found a way to criticize the government in Stalinist Russia and *totally get away with it*.

Who knows, you might even get that cushy professorship.

If you "criticize" society by telling it to keep doing exactly what it's doing only much much more so, society recognizes you as an ally and rewards you for being a "bold iconoclast" or "having brave and revolutionary new ideas" or whatever. It's only when you tell them something they *actually don't want to hear* that you get in trouble.

Western society has been moving gradually further to the left for the past several hundred years at least. It went from divine right of kings to constutitional monarchy to libertarian democracy to federal democracy to New Deal democracy through the civil rights movement to social democracy to ???. If you catch up to society as it's pushing leftward and say "Hey guys, I think we should go leftward even faster! Two times faster! No, *fifty* times faster!", society will call you a bold revolutionary iconoclast and give you a professorship.

If you start suggesting maybe it should switch directions and move the direction opposite the one the engine is pointed, *then* you might have a bad time.

Try it. Mention that you think we should undo something that's been done over the past century or two. Maybe reverse women's right to vote. Go back to sterilizing the disabled and feeble-minded. If you *really* need convincing, suggest re-implementing segregation, or how about slavery? See how far freedom of speech gets you.

In America, it will get you fired from your job and ostracized by nearly everyone. Depending on how loudly you do it, people may picket your house, or throw things at you, or commit violence against you which is then excused by the judiciary because obviously they were provoked. Despite the

iconic image of the dissident sent to Siberia, this is how the Soviets dealt with most of *their* iconoclasts too.

If you absolutely insist on imprisonment, you can always go to Europe, where there are more than enough "hate speech" laws on the book to satisfy your wishes. But a system of repression that doesn't involve obvious state violence is little different in effect than one that does. It's simply more efficient and harder to overthrow.

Reaction isn't a conspiracy theory; it's not suggesting there's a secret campaign for organized repression. To steal an example from the other side of the aisle, it's positing something more like patriarchy. Patriarchy doesn't have an actual Patriarch coordinating men in their efforts to keep down women. It's just that when lots of people share some really strong cultural norms, they manage to self-organize into a kind of immune system for rejecting new ideas. And Western society just happens to have a really strong progressivist immune system ready to gobble you up if you say anything insufficiently progressive.

And so the main difference between modern liberal democracy and older repressive societies is that older societies repressed things you liked, but modern liberal democracies only repress things you don't like. Having only things you don't like repressed looks from the inside a lot like there being no repression at all.

The good Catholic in medieval Spain doesn't feel repressed, even when the Inquisition drags away her neighbor. She feels like decent people have total freedom to worship whichever saint they want, total freedom to go to whatever cathedral they choose, total freedom to debate who the next bishop should be – oh, and thank goodness someone's around to deal with those

crazy people who are trying to damn the rest of us to Hell. We medieval Spaniards are way too smart to fall for the balance fallacy!

Wait, You Mean The Invisible Multi-Tentacled Monster That Has Taken Over All Our Information Sources Might Be Trying To *Mislead* Us?

Since you are a citizen of a repressive society, you should be extremely skeptical of all the information you get from schools, the media, and popular books on any topic related to the areas where active repression is occurring. That means at *least* politics, history, economics, race, and gender. You should be *especially* skeptical of any book that's praised as "a breath of fresh air" or "a good counter to the prevailing bias", as books that garner praise in the media are probably of the "We need fifty Stalins!" variety.

This is not nearly as paranoid as it sounds. Since race is the most taboo subject in our culture, it will also be the simplest example. Almost all of our hard data on race comes from sociology programs in universities – ie the most liberal departments in the most liberal institutions in the country. Most of these sociology departments have an explicit mission statement of existing to fight racism. Many sociologists studying race will tell you quite openly that they went into the field – which is not especially high-paying or prestigious – in order to help crusade against the evil of racism.

Imagine a Pfizer laboratory whose mission statement was to prove Pfizer drugs had no side effects, and whose staff all went into pharmacology specifically to help crusade against the evil of believing Pfizer's drugs have side effects. Imagine that this laboratory hands you their study showing that the

latest Pfizer drug has zero side effects, c'mon, trust us! Is there *any way* you're taking that drug?

We know that a lot of medical research, especially medical research by drug companies, turns up the wrong answer simply through the file-drawer effect. That is, studies that turn up an exciting result everyone wants to hear get published, and studies that turn up a disappointing result don't – either because the scientist never submits it to the journals, or because the journal doesn't want to publish it. If this happens *all the time* in medical research despite growing safeguards to prevent it, how often do you think it happens in sociological research?

Do you think the average sociologist selects the study design most likely to turn up evidence of racist beliefs being correct, or the study design most likely to turn up the opposite? If despite her best efforts a study does turn up evidence of racist beliefs being correct, do you think she's going to submit it to a major journal with her name on it for everyone to see? And if by some bizarre chance she does submit it, do you think the *International Journal Of We Hate Racism So We Publish Studies Proving How Dumb Racists Are* is going to cheerfully include it in their next edition?

And so when people triumphantly say "Modern science has completely disproven racism, there's not a shred of evidence in support of it", we should consider that exactly the same level of proof as the guy from 1900 who said "Modern science has completely proven racism, there's not a shred of evidence against it". The field is still just made of people pushing their own dogmatic opinions and calling them science; only the dogma has changed.

And although Reactionaries love to talk about race, in the end race is nothing more than a particularly strong and obvious taboo. There are taboos in history, too, and in economics, and in political science, and although they're less obvious and interesting they still mean you need this same skepticism when parsing results from these fields. "But every legitimate scientist disagrees with this particular Reactionary belief!" should be said with the same intonation as "But every legitimate archbishop disagrees with this particular heresy."

This is not intended as a proof that racism is correct, or even as the slightest shred of evidence for that hypothesis (although a lot of Reactionaries are, in fact, racist as heck). No doubt the Spanish Inquisition found a couple of real Satanists, and probably some genuine murderers and rapists got sent to Siberia. Sometimes, once in a blue moon, a government will even censor an idea that happens to be false. But it's still useful to know when something is being censored, so you don't actually think the absence of evidence for one side of the story is evidence of anything other than people on that side being smart enough to keep their mouths shut.

The Past Is A First World Country

Even so, isn't the evidence that modern society beats past societies kiiiind of overwhelming? We're richer, safer, healthier, better educated, freer, happier, more equal, more peaceful, and more humane. Reactionary responses to these claims might get grouped into three categories.

The first category is "Yes, obviously". Most countries do seem to have gotten about 100x wealthier since the year 1700. Disease rates have plummeted, and life expectancy has gone way up – albeit mostly due to changes in infant mortality. But this stands entirely explained by technology. So we're a

hundred times wealthier than in 1700. In what? Gold and diamonds? Maybe that has something to do with the fact that today we're digging our gold mines with one of these:

…and in 1700 they had to dig their gold mines with one of these:

Likewise, populations are healthier today because they can get computers to calculate precisely targeted radiation bursts that zap cancer while sparing healthy tissue, whereas in 1700 the pinnacle of medical technology was leeches.

This technology dividend appears even in unexpected places. The world is more peaceful today, but how much of that is the existence of global trade networks that make war unprofitable,

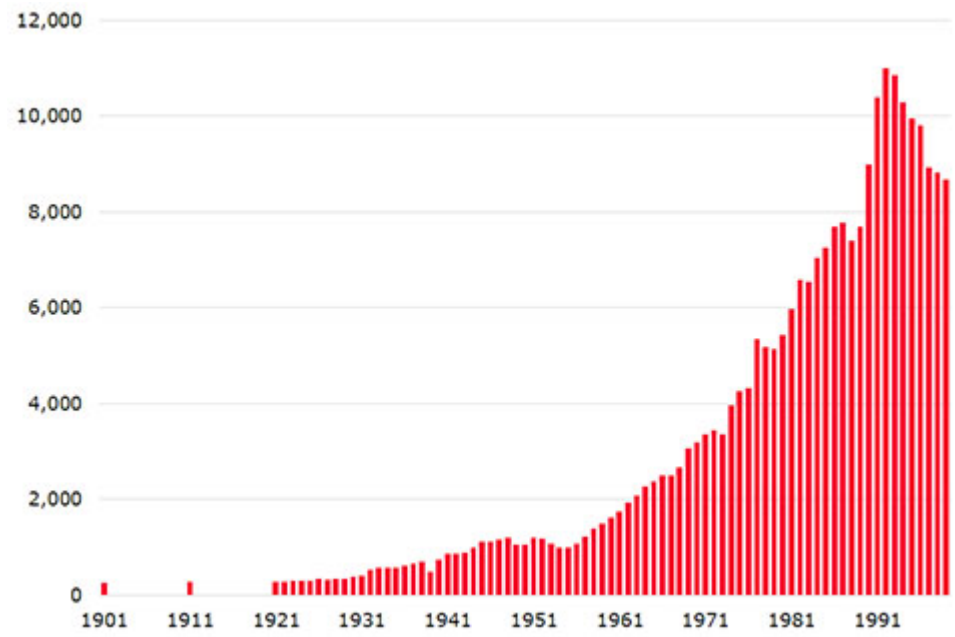
video reporting of every casualty that makes war unpopular, and nuclear and other weapons that make war unwinnable?

The second category is “oh really?”. Let’s take safety. This is one of Mencius Moldbug’s pet issues, and he likes to quote the following from an 1876 century text on criminology:

> Meanwhile, it may with little fear of contradiction be asserted that there never was, in any nation of which we have a history, a time in which life and property were so secure as they are at present in England. The sense of security is almost everywhere diffused, in town and country alike, and it is in marked contrast to the sense of insecurity which prevailed even at the beginning of the present century. There are, of course, in most great cities, some quarters of evil repute in which assault and robbery are now and again committed. There is perhaps to be found a lingering and flickering tradition of the old sanctuaries and similar resorts. But any man of average stature and strength may wander about on foot and alone, at any hour of the day or the night, through the greatest of all cities and its suburbs, along the high roads, and through unfrequented country lanes, and never have so much as the thought of danger thrust upon him, unless he goes out of his way to court it.

Moldbug then usually contrasts this with whatever recent news article has struck his fancy about entire inner-city neighborhoods where the police are terrified to go, teenagers being mowed down in crossfire among gangs, random daylight murders, and the all the other joys of life in a 21st century British ghetto.

Of course, the plural of anecdote is not data, but the British crime statistics seem to bear him out:



(recorded offenses per 100,000 people, from source)

If this is true, it is true *despite* technology. If crime rates have in fact multiplied by a factor of…well, it looks like at least 100x…this is true even though the country as a whole has gotten vastly richer, even though there are now CCTVs, DNA testing, police databases, heck, even fingerprinting hadn't been figured out yet in 1876.

This suggests that there was something inherent about Victorian society, politics, or government that made their Britain a safer place to live than modern progressive Britain.

Education is another example of something we're pretty sure we do better in. Now take a look at the 1899 entrance exam for Harvard. Remember, no calculators – they haven't been invented yet.

I got an SAT score well above that of the average Harvard student today (I still didn't get into Harvard, because I was a slacker in high school). But I couldn't even *begin* to take much of that test.

Okay, fine. Argue "Well, of course we don't value Latin and Greek and arithmetic and geometry and geography today, we value different things." So fine. Tell me what the heck you think our high school students are learning that's just as difficult and impressive as the stuff on that test that you don't expect the 19th century Harvard students who aced that exam knew two hundred times better (and don't say "the history of post-World War II Europe").

Do you honestly think the student body for whom that exam was a fair ability test would be befuddled by the *reading comprehension* questions that pass for entrance exams today? Or would it be more like "Excuse me, teacher, I'm afraid there's been a mistake. My exam paper is in English."

As a fun exercise, read through Wikipedia's list of multilingual presidents of the United States. We start with entries like this one:

> Thomas Jefferson read a number of different languages. In a letter to Philadelphia publisher Joseph Delaplaine on April 12, 1817, Jefferson claimed to read and write six languages: Greek, Latin, French, Italian, Spanish, and English. After his death, a number of other books, dictionaries, and grammar manuals in various languages were found in Jefferson's library, suggesting that he studied additional languages beyond those he spoke and wrote well. Among these were books in Arabic, Gaelic, and Welsh.

and this one:

> John Quincy Adams went to school in both France and the Netherlands, and spoke fluent French and conversational Dutch. Adams strove to improve his

> abilities in Dutch throughout his life, and at times translated a page of Dutch a day to help improve his mastery of the language. Official documents that he translated were sent to the Secretary of State of the United States, so that Adams' studies would serve a useful purpose as well. When his father appointed him United States Ambassador to Prussia, Adams dedicated himself to becoming proficient in German in order to give him the tools to strengthen relations between the two countries. He improved his skills by translating articles from German to English, and his studies made his diplomatic efforts more successful. In addition to the two languages he spoke fluently, he also studied Italian, though he admitted to making little progress in it since he had no one with whom to practice speaking and hearing the language. Adams also read Latin very well, translated a page a day of Latin text, and studied classical Greek in his spare time.

eventually proceeding to entries more like this one:

> George W. Bush speaks some amount of Spanish, and has delivered speeches in the language. His speeches in Spanish have been imperfect, with English dispersed throughout. Some pundits, like Molly Ivins, have pointedly questioned the extent to which he could speak the language, noting that he kept to similar phrasing in numerous appearances.

and this one:

> Barack Obama himself claims to speak no foreign languages. However, according to the President of Indonesia Susilo Bambang Yudhoyono, during a

> telephone conversation Obama was able to deliver a basic four-word question in “fluent Indonesian”, as well as mention the names for a few Indonesian food items. He also knows some Spanish, but admits to only knowing “15 words” and having a poor knowledge of the language.

A real Reactionary would no doubt point out that even old-timey US Presidents aren’t old-timey enough, and that we really should be looking at the British aristocracy, but this is left as an exercise for the reader.

It may be argued that yes, maybe their aristocracy was more educated than our upper-class, but we compensate for the imbalance by having education spread much more widely among the lower-classes. I endorse this position, as do, I’m sure, the hundreds of inner-city minority youth who are no doubt reading this blog post because of the massive interest in abstract political philosophy their schooling has successfully inspired in them.

Once again, today we have Wikipedia, the Internet, and as many cheap books as Amazon can supply us. Back in the old days they had to make do with whatever they could get from their local library. Even more troubling, today we start with a huge advantage – the Flynn Effect has made our average IQ 10 to 20 points higher than in 1900. Yet once again, even with our huge technological and biological head start, we are *still* doing worse than the Old Days, which suggest that here, too, the Old Days may have had some kind of social/political advantage.

So several of our claims of present superiority – wealth, health, peace, et cetera – have been found to be artifacts of higher technology levels. Several other claims – safety and education – have been found to be just plain wrong. That just

leaves a few political advantages – namely, that we are freer, less racist, less sexist, less jingoistic and more humane. And the introduction has already started poking holes in the whole "freedom" thing.

That leaves our progress in tolerance, equality, and humaneness. Are these victories as impressive as we think?

Every Time I Hear The Word "Revolver", I Reach For My Culture

[TRIGGER WARNING: This is the part with the racism]

One of the most solid results from social science has been large and persistent differences in outcomes across groups. Of note, these differences are highly correlated by goodness: some groups have what we would consider "good outcomes" in many different areas, and others have what we would consider "bad outcomes" in many different areas. Crime rate, drug use, teenage pregnancy, IQ, education level, median income, health, mental health, and whatever else you want to measure.

The best presentation of this result is The Spirit Level, even though the book *thinks* it's proving something completely different. But pretty much any study even vaguely in this field will show the same effect. This also seems to be the intuition behind our division of countries into "First World" and "Third World", and behind our division of races into "privileged" and "oppressed" (rather than "well, some races have good outcomes in some areas, but others have good outcomes in other areas, so it basically all balances out") I don't think this part should be very controversial. Let's call this mysterious quality "luck", in order to remain as agnostic as possible about the cause.

Three very broad categories of hypothesis have been proposed to explain luck differences among groups: the external, the cultural, and the biological.

The externalists claim that groups differ only because of the situations they find themselves in. Sometimes these situations are natural. Jared Diamond makes a cogent case for the naturalist externalist hypothesis in *Guns, Germs, and Steel*. The Chinese found themselves on fertile agricultural land with lots of animals and plants to domesticate and lots of trade routes to learn new ideas from. The New Guinea natives found themselves in a dense jungle without many good plants or animals and totally cut off from foreign contact. Therefore, the Chinese developed a powerful civilization and the New Guineans became a footnote to history.

But in modern times, externalists tend to focus more on external *human* conditions like colonialism and oppression. White people are lucky not because of any inherent virtue, but because they had a head start and numerical advantage and used this to give themselves privileges which they deny to other social groups. Black people are unlucky not because of any inherent flaw, but because they happened to be stuck around white people who are doing everything they can to oppress them and keep them down. This is true both within societies, where unlucky races are disprivileged by racism, and across societies, where unlucky countries suffer the ravages of colonialism.

The culturalists claim that luck is based on the set of implicit traditions and beliefs held by different groups. The Chinese excelled not only because of their fertile landscape, but because their civilization valued scholarship, wealth accumulation, and nonviolence. The New Guineans must have had less useful values, maybe ones that demanded strict

conformity with ancient tradition, or promoted violence, or discouraged cooperation.

Like the externalists, they trace this forward to the present, saying that the values that served the Chinese so well in building Chinese civilization are the same ones that keep China strong today and the ones that make Chinese immigrants successful in countries like Malaysia and the USA. On the other hand, New Guinea continues to be impoverished and although I've never heard of any New Guinean immigrants I would not expect them to do very well.

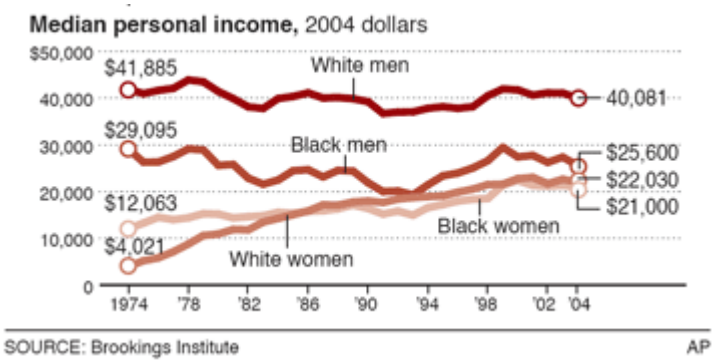
The biologicalists, for whom I cannot think of a less awkward term, are probably the most notorious and require the least explanation. They are most famous for attributing between-group luck differences to genetic factors, but there are certainly more subtle theories. One of the most interesting is parasite load, the idea that areas with greater parasites make people's bodies spend more energy fighting them off, leading to less energy for full neurological development. It's hard to extend this to deal with group differences in a single area (for example between-race differences in the USA) but some people have certainly made valiant attempts. Nevertheless, it's probably fair enough to just think of the biologicalists as "more or less racists".

So who is right?

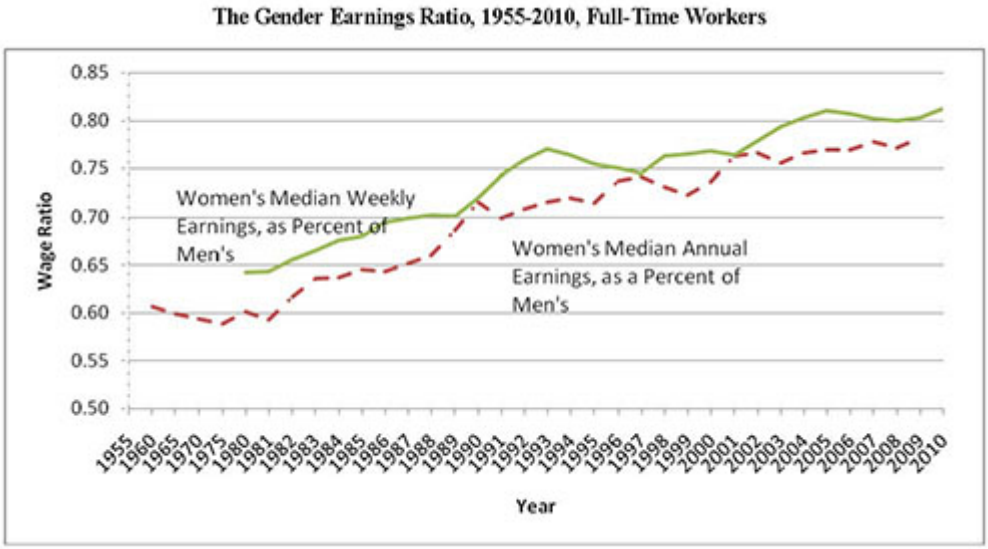
A decent amount of political wrangling over the years seems to involve a conflict between the conservatives – who are some vague mix of the culturalist and biologicalist position – and the liberals, who have embraced the externalist position with gusto.

But the externalist position is deeply flawed. This blog has already cited this graph to make a different point, but now that

we have our Reactionary Hat on, let's try it again:



Here's the black-white income gap over time from 1974 to (almost) the present. Over those years, white oppression of black people has decreased drastically. It is not gone. But it has decreased. Yet the income gap stays exactly the same. Compare this to another example of an oppressed group suddenly becoming less oppressed:



Over the same period, the decrease in male oppression of women has resulted in an obvious and continuing rise in women's incomes. This suggests that the externalist hypothesis of women's poor incomes was at least partly correct. But an apparent corollary is that it casts doubt on the externalist hypothesis of racial income gaps.

And, in fact, not all races have a racial income gap, and not all those who do have it in the direction an externalist theory would predict. Jews and Asians faced astounding levels of discrimination when they first came to the United States, but both groups recovered quickly and both now do significantly better than average white Americans. Although the idea of a "Jewish conspiracy" is rightly mocked as anti-Semitic and stupid, it is only bringing the externalist hypothesis – that differences in the success of different races must always be due to oppression – to their natural conclusion.

In fact, Jews and Chinese are interesting in that both groups are widely scattered, both groups often find themselves in very hostile countries, and yet both groups are usually more successful than the native population wherever they go (income and education statistics available upon request). Whether it is Chinese in Malaysia or Jews in France, they seem to do unusually well for themselves despite the constant discrimination. If this is an experiment to distinguish between culturalist and externalist positions, it is a very well replicated one.

This difference in the success of immigrant groups is often closely correlated with the success of the countries they come from. Japan is very rich and advanced, Europe quite rich and advanced, Latin America not so rich or advanced, and Africa least rich and advanced of all. And in fact we find that Japanese-Americans do better than European-Americans do better than Latin-Americans do better than African-Americans. It is pretty amazing that white people manage to modulate their oppression in quite this precise a way, especially when it includes oppressing themselves.

And much of the difference between groups is in areas one would expect to be resistant to oppression. Unlucky groups

tend to have higher teenage pregnancy rates, more drug use, and greater intra-group violence, *even when comparing similar economic strata*. That is, if we focus on Chinese-Americans who earn $60,000/year and African-Americans who earn $60,000/year, the Chinese will have markedly better outcomes (I've seen this study done in education, but I expect it would transfer). Sampling from the same economic stratum screens off effects from impoverished starting conditions or living in bad neighborhoods, and it's hard (though of course not impossible) to figure out other ways an oppressive majority could create differential school attendance in these groups.

So luck differences are sometimes in favor of oppressed minorities, do not decrease when a minority becomes less oppressed, correlate closely across societies with widely varying amounts of oppression, and operate in areas where oppression doesn't provide a plausible mechanism. The externalist hypothesis as a collection of natural factors a la Jared Diamond may have merit, but as an oppression-based explanation for modern-day group differences, it fails miserably.

I don't want to dwell on the biological hypothesis too much, because it sort of creeps me out even in a "let me clearly explain a hypothesis I disagree with" way. I will mention that it leaves a lot unexplained, in that many of the "groups" that have such glaring luck differences are not biological groups at all, but rather religious groups such as the Mormons and the Sikhs, both of whom have strikingly different outcomes than the populations they originated from. Even many groups that are biologically different just aren't different *enough* – the English and Irish have strikingly different luck, but attributing that to differences between which *exact* tiny little branch of the Indo-European tree they came from seems like a terrible

explanation (although Konkvistador disagrees with me on this one).

Nevertheless, the people who dismiss the biological hypothesis as obviously stupid and totally discredited (by which I mean everyone) are doing it a disservice. For a sympathetic and extraordinarily impressive defense of the biological hypothesis I recommend this unpublished (and unpublishable) review article. I will add that I am *extremely* interested in comprehensive takedowns of that article (preferably a full fisking) and that if you have any counterevidence to it at all you should post it in the comments and I will be eternally grateful.

But for now I'm just going to say let's assume by fiat that the biologicalist hypothesis is false, because even with my Reactionary Hat on I find the culturalist hypothesis much more interesting.

The culturalist hypothesis avoids the pitfalls of both the externalist and biological explanations. Unlike the externalists, it can explain why some minority groups are so successful and why group success correlates across societies and immigrant populations. And unlike the biologicalists, it can explain the striking differences between biologically similar groups like the Mormons and the non-Mormon Americans, or the Sikhs and the non-Sikh Indians.

It can also explain some other lingering mysteries, like why a country that's put so much work into keeping black people down would then turn around and elect a black president. Obama was born to an African father and a white mother, raised in Indonesia, and then grew up in Hawaii. At no point did he have much contact with African-American culture, and

so a culturalist wouldn't expect his life outcomes to be correlated with those of other African-Americans.

Best of all, despite what the average progressive would tell you the culturalist position isn't really *that* racist. It's a bit like the externalist position in attributing groups' luck to initial conditions, except instead of those initial conditions being how fertile their land is or who's oppressing them, it's what memeplexes they happened to end out with. Change the memeplexes and you can make a New Guinean population achieve Chinese-level outcomes – or vice versa.

The Other Chinese Room Experiment

Assuming we tentatively accept the culturalist hypothesis, what policies does it suggest?

Well, the plan mentioned in the last paragraph of the last section – throw Chinese memes at the people of New Guinea until they achieve Chinese-style outcomes – higher income, less teenage pregnancy, lower crime rates. It doesn't seem like a bad idea. You could try exposing them to Chinese people and the Chinese way of life until some of it stuck. This seems like a good strategy for China, a country whose many problems definitely do not include "a shortage of Chinese people".

On the other hand, in somewhere more like America, one could be forgiven for immediately rounding this off to some kind of dictatorial brainwashing policy of stealing New Guinean infants away from their homes and locking them in some horrible orphanage run by Chinese people who beat them every time they try to identify with their family or native culture until eventually they absorb Chinese culture through osmosis. This sounds bad.

Luckily, although we don't have quite as many Chinese people as China, we still have a majority culture whose outcomes are

almost as good as China's and which, as has been mentioned before, permeates every facet of life and every information source like a giant metastasizing thousand-tentacled monster. So in theory, all we need to do is wait for the unstoppable monster to get them.

This strategy, with the octopoid abomination metaphor replaced with a melting pot metaphor for better branding, has been America's strategy for most of the past few centuries – assimilation. It worked for the Irish, who were once viewed with as much racism as any Hispanic or Arab is today. It worked for the Italians, who were once thought of as creepy Papist semi-retarded mafia goons until everyone decided no, they were indistinguishable from everyone else. It worked for the fourth and fifth generation Asians, at least here in suburban California, where they're considered about as "exotic" as the average Irishman. It certainly worked for the Jews, where there are some people of Jewish descent who aren't even *aware* of it until they trace their family history back. And it should be able to work for everyone else. Why isn't it?

The Reactionary's answer to this is the same as the Reactionary's answer to almost everything: because of those darned progressives!

Sometime in the latter half of this century, it became a point of political pride to help minorities resist "cultural imperialism" and the Eurocentric norms that they should feel any pressure to assimilate. Moved by this ideology, the government did everything it could to help minorities avoid assimilation and to shame and thwart anyone trying to get them to assimilate.

There's a story – I've lost the original, but it might have been in Moldbug – about a state noticing that black children were getting lower test scores. It decided, as progressivists do, that

the problem was that many of the classes were taught by white teachers, and that probably this meant the black children couldn't relate to them and were feeling oppressed. So they sent the white teachers off to whiter areas and hiring only black teachers for the black schools, and – sure enough – test scores plummeted further.

California had a sort of similar problem when I was growing up. Most schools were required to teach our large Hispanic immigrant population using bilingual education – that is, teaching them in their native Spanish until they were ready to learn English. The "ready to learn English" tended not to happen, and some people proposed that bilingual education be scrapped. There was a *huge* ruckus where the people in favor of this change were accused of being vile racists who hated Mexicans and wanted to destroy Mexican culture. Thanks to California's colorful proposition system, it passed anyway. And sure enough, as soon as the Hispanics started getting integrated with everyone else and taught in English, test scores went way up.

But this is a rare victory, and we are still very much in "try to prevent assimilation mode". I went to elementary school just as the "melting pot" metaphor was being phased out in favor of the more politically correct "salad bowl" one – in a melting pot, everyone comes together and becomes alike, but in a salad bowl, everything comes together but stays different, and that's fine.

One externalist argument why minorities sometimes do poorly in school is the fear of "acting white" – that their peers tell them that academic achievement is a form of "acting white" by which they betray their cultural heritage. Unfortunately, we seem to be promoting this on a social level, telling people that assimilating and picking up the best features of majority

culture are “acting white”. If the majority culture has useful memes that help protect people against school dropout, crime, and other bad life outcomes, that is a really bad thing to do.

So let’s go back to the nightmare scenario with which we started this section – of children being seized from their homes and locked in a room with Chinese people. Is this sort of dystopia the inevitable result of trying to use culturalist theories to equalize group outcomes?

No. There is a proverb beloved of many Reactionaries: “If you find yourself in a hole, stop digging.” We could make great strides in solving inequality merely by *ceasing to exert deliberate effort to make things worse*. The progressive campaign to demonize assimilation and make it taboo to even talk about some cultures being better adapted than others prevents the natural solution to inequality which worked for the Irish and the Asians and the Jews from working for the minorities of today. If we would *just stop digging the hole deeper* in order to make ourselves feel superior to our ancestors, we’d have gone a lot of the way – maybe not all of the way, but a lot of it – toward solving the problem.

On Second Thought, Keep Your Tired And Poor To Yourself

Immigration doesn’t have to be a problem. In a healthy society, immigrants will be encouraged to assimilate to the majority culture, and after a brief period of disorientation will be just as successful and well-adapted as everyone else.

But in an unhealthy society like ours that makes assimilation impossible, a culturalist will be very worried about immigration.

Let’s imagine an idyllic socialist utopia with a population of 100,000. In Utopia, everyone eats healthy organic food,

respects the environment and one another, lives in harmony with people of other races, and is completely non-violent. One day, the Prime Minister decides to open up immigration to Americans and discourage them from assimilating.

50,000 Americans come in and move into a part of Utopia that quickly becomes known as Americatown. They bring their guns, their McDonalds, their megachurches, and their racism.

Soon, some Utopians find their family members dying in the crossfire between American street gangs. The megachurches convert a large portion of the Utopians to evangelical Christianity, and it becomes very difficult to get abortions without being harassed and belittled. Black and homosexual Utopians find themselves the target of American hatred, and worse, some young Utopians begin to get affected by American ideas and treat them the same way. American litter fills the previously pristine streets, and Americans find some loopholes in the water quality laws and start dumping industrial waste into the rivers.

By the time society has settled down, we have a society which is maybe partway between Utopia and America. The Americans are probably influenced by Utopian ideas and not quite as bad as their cousins who reminded behind in the States, but the Utopians are no longer as idyllic as their Utopian forefathers, and have inherited some of America's problems.

Would it be *racist* for a Utopian to say "Man, I wish we had never let the Americans in?" Would it be *hateful* to suggest that the borders be closed before even more Americans can enter?

If you are a culturalist, no. Utopian culture is better, at least by Utopian standards, than American culture. Although other

cultures can often contribute to enrich your own, there is no law of nature saying that only the good parts of other cultures will transfer over and that no other culture can be worse than yours in any way. The Americans were clearly worse than the Utopians, and it was dumb of the Utopians to let so many Americans in without any safeguards.

Likewise, there are countries that are worse than America. Tribal Afghanistan seems like a pretty good example. Pretty much everything about tribal Afghanistan is horrible. Their culture treats women as property, enforces sharia law, and contains honor killings as a fact of life. They tend to kill apostate Muslims and non-Muslims a lot. Not all members of Afghan tribes endorse these things, but the average Afghan tribesperson is much more likely to endorse them than the average American. If we import a bunch of Afghan tribesmen, their culture is likely to make America a worse place in the same way that American culture makes Utopia a worse place.

But it's actually much worse than this. We are a democracy. Anyone who moves here and gains citizenship eventually gets the right to vote. People with values different from ours vote for people and laws different from those we would vote for. Progressives have traditionally viewed any opposition to this as anti-immigrant and racist – and, by total coincidence, most other countries, and therefore most immigrants, are progressive.

Imagine a country called Conservia, a sprawling empire of a billion people that has a fifth-dimensional hyperborder with America. The Conservians are all evangelical Christians who hate abortion, hate gays, hate evolution, and believe all government programs should be cut.

Every year, hundreds of thousands of Conservians hop the hyperborder fence and enter America, and sympathetic presidents then pass amnesty laws granting them citizenship. As a result, the area you live – or let's use Berkeley, the area I live – gradually becomes more conservative. First the abortion clinics disappear, as Conservian protesters start harassing them out of business and a government that must increasingly pander to Conservians doesn't stop them. Then gay people stop coming out of the closet, as Conservian restaurants and businesses refuse to serve them and angry Conservian writers and journalists create an anti-gay climate. Conservians vote 90% Republican in elections, so between them and the area's native-born conservatives the Republicans easily get a majority and begin defunding public parks, libraries, and schools. Also, Conservians have one pet issue which they promote even more intently than the destruction of secular science – that *all Conservians illegally in the United States must be granted voting rights, and that no one should ever block more Conservians from coming to the US*.

Is this fair to the native Berkeleyans? It doesn't seem that way to me. And what if 10 million Conservians move into America? That's not an outrageous number – there are more Mexican immigrants than that. But it would be enough to have thrown every single Presidential election of the past fifty years to the Republicans – there has never been a Democratic candidate since LBJ who has won the native population by enough of a margin to outweight the votes of ten million Conservians.

But isn't this incredibly racist and unrealistic? An entire nation of people whose votes skew 90% Republican? No. African-Americans' votes have historically been around 90% Democratic (93% in the last election). Latinos went over 70%

Democratic in the last election. For comparison, white people were about 60% Republicans. If there had been no Mexican immigration to the United States over the past few decades, Romney would probaby have won the last election.

Is it wrong for a liberal citizen of Berkeley in 2013 to want to close the hyperborder with Conservia so that California doesn't become part of the Bible Belt and Republicans don't get guaranteed presidencies forever? Would that citizen be racist for even considering this? If not, then pity the poor conservative, who is actually in this exact situation right now.

(a real Reactionary would hasten to add this is more proof that progressives control everything. Because immigration favors progressivism, any opposition to it is racist, but the second we discover the hyperborder with Conservia, the establishment will figure out some reason why *allowing* immigration is racist. Maybe they can call it "inverse colonialism" or something.)

None of this is an argument against immigration. It's an argument against immigration by groups with bad Luck and with noticeably different values than the average American. Let any Japanese person who wants move over. Same with the Russians, and the Jews, and the Indians. Heck, it's not even like it's saying no Afghans – if they swear on a stack of Korans that they're going to try to learn English and not do any honor killings, they could qualify as well.

The United States used to have a policy sort of like this. It was called the Immigration Act of 1924. Its actual specifics were dumb, because it banned for example Asians and Jews, but the principle behind it – groups with good outcomes and who are a good match for our values can immigrate as much as they want, everyone else has a slightly harder time – seems broadly

wise. So of course progressives attacked it as racist and Worse Than Hitler and it got repealed in favor of the current policy: *everyone* has a really hard time immigrating but if anyone sneaks over the border under cover of darkness we grant them citizenship anyway because not doing that would be mean.

Once again, coming up with a fair and rational immigration policy wouldn't require some incredibly interventionist act of state control. It would just require that we notice the hole we've been deliberately sticking ourselves in and *stop digging*.

Imperialism Strikes Back

In an externalist/progressive worldview, the best way to help disadvantaged minorities is to eliminate the influence of more privileged majority groups. In a culturalist/Reactionary worldview, the best way to help disadvantaged minorities is to try to maximize the influence of more privileged majority groups. This suggests re-examining colonialism. But first, a thought experiment.

Suppose you are going to be reincarnated as a black person (if you are already black, as a different black person). You may choose which country you will be born in; the rest is up to Fate. What country do you choose?

The top of my list would be Britain, with similar countries like Canada and America close behind. But what if you could only choose among majority-black African countries?

Several come to my mind as comparatively liveable. Kenya. Tanzania. Botswana. South Africa. Namibia (is your list similar?) And one thing these places all have in common was being heavily, *heavily* colonized by the British.

We compare the sole African country that was never colonized, Ethiopia. Ethiopia has become a byword for

senseless suffering thanks to its coups, wars, genocides, and especially famines. This seems like counter-evidence to the "colonialism is the root of all evil" hypothesis.

Yes, colonization had some horrible episodes. Anyone who tries to say King Leopold II was anything less than one of the worst people who ever lived has zero right to be taken seriously. On the other hand, eventually the Belgian people got outraged enough to take it away from Leopold, after which there follows a fifty year period that was the only time in history when the Congo was actually a kind of nice place. Mencius Moldbug likes to link to a Time magazine article from the 1950s praising the peace and prosperity of the Congo as a model colony. Then in 1960 it became independent, and I don't know what happens next because the series of civil wars and genocides and corrupt warlords after that are so horrible that I can't even read all the way through the articles about them. Seriously, not necessarily in numbers but in sheer graphic brutality it is worse than the Holocaust, the Inquisition, and Mao combined and you *do not want to know* what makes me say this.

So yes, Leopold II is one of history's great villains, but once he was taken off the scene colonial Congo improved markedly. And any attempt to attribute the nightmare that is the modern Congo to colonialism has to cope with the historical fact that the post-Leopold colonial Congo was actually pretty nice until it was decolonized at which point it immediately went to hell.

So the theory that colonialism is the source of all problems has to contend with the observation that heavily colonized countries are the most liveable, the sole never-colonized country is among the least liveable, and countries' liveability plummeted drastically as soon as colonialism stopped.

But let's stop picking on Africans. Suppose you are going to be reincarnated as a person of Middle Eastern descent (I would have said "Arab", but then we would get into the whole 'most Middle Easterners are not Arabs' debate). Once again, you can choose your country. Where do you go?

Once again, Britain, US, or somewhere of that ilk sound like your best choices.

Okay, once again we're ruling that out. You've got to go somewhere in the Middle East.

Your best choice is one of those tiny emirates where everyone is a relative of the emir and gets lots of oil money and is super-rich: I would go with Qatar. Let's rule them out too.

Your next-best choice is Israel.

Yes, Israel. Note that I am *not* saying the Occupied Palestinian Territories; that would be just as bad a choice as you expect. I'm saying Israel, where 20% of the population is Arab, and about 16% Muslim.

Israeli Arabs earn on average about $6750 per year. Compare this to conditions in Israel's Arab neighbors. In Egypt, average earnings are $6200; in Jordan, $5900; in Syria, only $5000.

Aside from the economics, there are other advantages. If you happen to be Muslim, you will have a *heck* of a lot easier time practicing your religion freely in Israel than in some Middle Eastern country where you follow the wrong sect of Islam. You'll be allowed to vote for your government, something you can't do in monarchical Jordan or war-torn Syria, and which Egypt is currently having, er, severe issues around. You can even criticize the government as much as you want (empirically quite a lot), a right Syrian and Egyptian Arabs are currently dying for. Finally, you get the benefit of living in a

clean, safe, developed country with good health care and free education for all.

I'm not saying that Israeli Arabs aren't discriminated against or have it as good as Israeli Jews. I'm just saying they have it better than Arabs in most other countries. Once again, we find that colonialism, supposed to be the root of all evil, is actually preferable to non-colonialism in most easily measurable ways.

It may be the case that pre-colonial societies were better than either colonial or post-colonial societies. I actually suspect this is true, in a weird Comanche Indians are better than all of us sort of sense. But "pre-colonial" isn't a choice nowadays. Nowadays it's "how much influence do we want the better parts of the West to have over countries that have already enthusiastically absorbed the worst parts of the West?" Whatever I may feel about the Safavid Dynasty, I would at least rather be born in Afghanistan-post-American-takeover than Afghanistan-pre-American-takeover.

So does this mean some sort of nightmarish "invade every country in the world, kill their leadership, and replace them with Americans, for their own good" type scenario?

Once again, no. Look at China. They've been quietly colonizing Africa for a decade now, and the continent has never been doing better. And by "colonizing", I mean "investing in", with probably some sketchy currying of influence and lobbying and property-gathering going on on the side. It's been great for China, it's been a hugely successful injection of money and technology into Africa, and they probably couldn't have come up with a better humanitarian intervention if they had been trying.

Why hasn't the West done it? Because every time an idea like that has been mooted, the progressives have shot it down with

“You neo-colonialist! You’re worse than King Leopold II, who was himself worse than Hitler! By the transitive property, you are *worse than Hitler*!”

No one needs to go about invading anyone else or killing their government. But if you find yourself in a hole, *stop digging*.

The Uncanny Valley Of Dictatorship

I kind of skimmed over the Palestinian Territories in the last section. They are, indeed, a terrible dehumanizing place and the treatment of their citizens is an atrocity that blemishes a world which allows it to continue. Is this a strike against colonialism?

Any 19th century European aristocrat looking at the Palestinian Territories would note that Israel is being a *terrible colonizer*, not in a moral sense but in a purely observational sense. It’s not getting any money or resources out of its colony at all! It’s letting people totally just protest it and get away with it! They’ve even handed most of it over to a government of natives! Queen Victoria would *not* be amused.

Suppose a psychopath became Prime Minister of Israel (yes, obvious joke is obvious). He declares: “Today we are annexing the Palestinian territories. All Palestinians become Israeli residents with most of the rights of citizens except they can’t vote. If anyone speaks out against Israel, we’ll shoot them. If anyone commits a crime, we’ll shoot them.” What would happen?

Well, first, a lot of people would get shot. After that? The Palestinians would be in about the same position as Israeli Arabs are today, except without the right to vote, plus they get shot if they protest. This is vastly better than the position they’re in now, and better than the position of say the people

of Syria who are poorer, *also* lack the right to vote, and *also* empirically get shot if they protest.

No more worries about roadblocks. No more worries about passports. No more worries about sanctions. No more worries about economic depression. The only worry is getting shot, and you can avoid that by never speaking out against Israel. Optimal? Probably not. A heck of a lot better than what the Palestinians have today? Seems possible.

It seems like there's an uncanny valley of dictatorship. Having no dictator at all, the way it is here in America, is very good. Having a really really dictatorial dictator who controls everything, like the czar or this hypothetical Israeli psychopath, kinda sucks but it's peaceful and you know exactly where you stand. Being somewhere in the middle, where it's dictatorial enough to hurt, but not dictatorial enough for the dictator to feel secure enough to mostly leave you alone except when he wants something, is worse than either extreme.

Mencius Moldbug uses the fable of Fnargl, an omnipotent and invulnerable alien who becomes dictator of Earth. Fnargl is an old-fashioned greedy colonizer: he just wants to exploit Earth for as much gold as possible. He considers turning humans into slaves to work in gold mines, except some would have to be a special class of geologist slaves to plan the gold mines, and there would have to be other slaves to grow food to support the first two classes of slaves, and other slaves to be managers to coordinate all these other slaves, and so on. Eventually he realizes this is kind of dumb and there's already a perfectly good economy. So he levies a 20% tax on every transaction (higher might hurt the economy) and uses the money to buy gold. Aside from this he just hangs out.

Fnargl has no reason to ban free speech: let people plot against him. He's omnipotent and invulnerable; it's not going to work. Banning free speech would just force him to spend money on jackbooted thugs which he could otherwise be spending on precious, precious gold. He has no reason to torture dissidents. What are they going to do if left unmolested? *Overthrow* him?

Moldbug claims that Fnargl's government would not only be better than that of a less powerful human dictator like Mao, but that it would be *literally better than the government we have today*. Many real countries *do* restrict free speech or torture dissidents. And if you're a libertarian, Fnargl's "if it doesn't disrupt gold production, I'm okay with it" line is a dream come true.

So if the Israelis want to improve the Palestinian Territories' plight, they can do one of two things. First, they can grant it full independence. Second, they do exactly the opposite: can take away all of its independence and go full Fnargl.

We already know Israel doesn't want to just grant full independence, which leaves "problem continues forever" or "crazy psychopath alien solution". Could the latter really work?

Well, no. Why not? Because the Palestinians would probably freak out and start protesting *en masse* and the Israelis would have to shoot all of them and that would be horrible.

But it's worth noting this is not just a natural state of the world. The British successfully colonized Palestine for several decades. They certainly tried the Fnargl approach: "No way you're getting independence, so just sit here and deal with it or we shoot you." It worked pretty well then. I would hazard a guess to say the average Palestinian did much better under

British rule than they're doing now. So why wouldn't it work again?

In a word, progressivism. For fifty years, progressives have been telling the colonized people of the world "If anyone colonizes you, this is the worst thing in the world, and if you have any pride in yourself you must start a rebellion, even a futile rebellion, immediately." This was non-obvious to people a hundred years ago, which is why people rarely did it. It was only after progressivism basically told colonized peoples "You're not revolting yet? What are you, *chicken*?" that the modern difficulties in colonialism took hold. And it's only after progressivism gained clout in the countries that rule foreign policy that it became politically impossible for a less progressive country to try colonialism.

If not for progressivism, Israel would have been able to peacefully annex the Palestinian territories as a colony with no more of a humanitarian crisis than Britain annexing New Zealand or somewhere. Everything would have been solved and everyone could have gone home in time for tea.

Once again, the problem with these holes is that we *keep digging them*. Maybe if we'd stop, there wouldn't be so many holes anymore.

Humane, All Too Humane

There seem to be similar uncanny valley effects in the criminal justice system and in war.

Modern countries pride themselves on their humane treatment of prisoners. And by "humane", I mean "lock them up in a horrible and psychologically traumatizing concrete jail for ten years of being beaten and raped and degraded, sometimes barely even seeing the sun or a green plant for that entire time, then put it on their permanent record so they can never get a

good job or interact with normal people ever again when they come out."

Compare this to what "inhumane" countries that were still into "cruel and unusual punishment" would do for the same crime. A couple of lashes with the whip, then you're on your way.

Reader. You have just been convicted of grand theft auto (the crime, not the game). You're innocent, but the prosecutor was very good at her job and you've used up all your appeals and you're just going to have to accept the punishment. The judge gives you two options:

1) Five years in prison
2) Fifty strokes of the lash

Like everyone else except a few very interesting people who help provide erotic fantasies for the rest of us, I don't like being whipped. But I would choose (2) in a *fraction of a heartbeat*.

And aside from being better for me, it would be better for society as well. We know that people who spend time in prison are both more likely to stay criminals in the future and better at being criminals. And each year in jail costs the State $50,000; more than it would cost to give a kid a year's free tuition at Harvard. Cutting the prison system in half would free up approximately enough money to give free college tuition to all students at the best school they can get into.

But of course we don't do that. We stick with the prisons and the rape and the kids who go work at McDonalds because they can't afford college. Why? *Progressives!*

If we were to try to replace prison with some kind of corporal punishment, progressives would freak out and say we were cruel and inhumane. Since the prison population is

disproportionately minority, they would probably get to use their favorite word-beginning-with-"R", and allusions would be made to plantation owners who used to whip slaves. In fact, progressives would come up with some reason to oppose even giving criminals the *option* of corporal punishment (an option most would certainly take) and any politician insufficiently progressive to even recommend it would no doubt be in for some public flagellation himself, albeit of a less literal kind.

So once again, we have an uncanny valley. Being very nice to prisoners is humane and effective (Norway seems to be trying ths with some success), but we're not going to do it because we're dumb and it's probably too expensive anyway. Being very strict to prisoners is humane and effective – the corporal punishment option. But being somewhere in the fuzzy middle is cruel to the prisoners and incredibly destructive to society – and it's the only route the progressives will allow us to take.

Some Reactionaries have tried to apply the same argument to warfare. Suppose that during the Vietnam War, we had nuked Hanoi. What would have happened?

Okay, fine. The Russians would have nuked us and everyone in the world would have died. Bad example. But suppose the Russians were out of the way. Wouldn't nuking Hanoi be a massive atrocity?

Yes. But compare it to the alternative. Nuking Hiroshima killed about 150,000 people. The Vietnam War killed about 3 million. The latter also had a much greater range of non-death effects, from people being raped and tortured and starved to tens of thousands ending up with post-traumatic stress disorder and countless lives being disrupted. If nuking Hanoi would have been an alternative to the Vietnam War, it would have been a *really really good* alternative.

Most of the countries America invades know they can't defeat the US military long-term. Their victory condtion is helping US progressives bill the war as an atrocity and get the troops sent home. So the enemy's incentive is to make the war drag on as long as possible and contain as many atrocities as possible. It's not too hard to make the war drag on, because they can always just hide among civilians and be relatively confident the US is too humane to risk smoking them out. And it's never too hard to commit atrocities. So they happily follow their incentives, and the progressives in the US happily hold up their side of the deal by agitating for the troops to be sent home, which they eventually are.

Compare this to the style of warfare in colonial days. "This is our country now, we're not leaving, we don't really care about atrocities, and we don't really care how many civilians we end up killing." It sounds incredibly ugly, but of colonial Britain or very-insistently-non-colonial USA, guess which one ended up pacifying Iraq after three months with only about 6,000 casualties, and guess which one took five years to re-establish a semblance of order and killed about 100,000 people in the process?

Once again we see an uncanny valley effect. Leaving Iraq alone completely would have been a reasonable humanitarian choice. Using utterly overwhelming force to pacify Iraq by any means necessary would have briefly been very ugly, but our enemies would have folded quickly and with a few assumptions this could also have been a reasonable humanitarian choice. But a wishy-washy half-hearted attempt to pacify Iraq that left the country in a state of low-grade poorly-defined war for nearly a decade was neither reasonable nor humanitarian.

Once again, the solution isn't some drastic nightmare scenario where all prisoners are tortured and all wars are fought with sarin nerve gas. It's that if prisoners *prefer* corporal punishment, progressives don't call "racism!" or "atrocity!" so loudly that it becomes politically impossible to give them what they want. Once again, all we have to do is *stop digging*.

Gender! And Now That I Have Your Attention, Let's Talk About Sex

So the two things Reactionaries like to complain about all the time are race and sex, and since we have *more* then gone overboard with our lengthy diversion into race, we might as well take a quick look at sex.

As far as I know, even the Reactionaries who are really into biological differences between races don't claim that women are intellectually inferior to men. I don't even think they necessarily believe there are biological differences between the two groups. And yet they are not really huge fans of feminism. Why?

Let's start with some studies comparing gender roles and different outcomes.

Surveys of women show that they were on average happier fifty years ago than they are today. In fact, in the 1950s, women generally self-reported higher happiness than men; today, men report significantly higher happiness than women. So the history of the past fifty years – a history of more and more progressive attitudes toward gender – have been a history of women gradually becoming worse and worse off relative to their husbands and male friends.

This doesn't *necessarily* condemn progressivism, but as the ancient proverb goes, it sure waggles its eyebrows

suggestively and gestures furtively while mouthing ‘look over there’.

To confirm, we would want to look within a single moment in time: that is, are feminist women with progressive gender roles *today* less happy than their traditionalist peers? The answer appears to be yes.

Amusingly, because we *do* still live in a society where these things couldn’t be published unless someone took a progressivist tack, the New York Times article quoted above ends by saying the *real* problem is that men are jerks who don’t do their share of the housework.

But when we actually study this, we find that progressive marriages in which men and women split housework equally are 50% more likely to end in divorce than traditional marriages where the women mostly take care of it. The same is true of working outside the home: progressive marriages where both partners work are more likely to end in divorce than traditional marriages where the man works and the woman stays home.

Maybe this is just because the same people who are progressive enough to defy traditional gender roles are also the same people who are progressive enough not to think divorce is a sin? But this seems unlikely: in general religious people get divorced *more* than the irreligious. And since I did promise we’d be talking about sex, consider the studies showing people in traditional marriages have better sex lives than their feminist and progressive friends. This doesn’t seem like something that could easily be explained merely by religion, unless religion has gotten *way* cooler since the last time I attended synagogue.

So why is this? I have heard some reactionaries say that although there are not intellectual differences between men and women, there are emotional differences, and that women are (either for biological or cultural reasons) more "submissive" to men's "dominant" – and a quick search of the BDSM community seems to both to validate the general rule and to showcase some very striking exceptions.

But my money would be on a simpler hypothesis. Every marriage involves conflict. The traditional concept of gender contains two roles that are divided in a time-tested way to minimize conflict as much as possible. In a perfect-spherical-cow sense, either the husband or the wife could step into either role, and it would still work just as well. But since men have been socialized for one role since childhood, and women socialized for the other role, it seems that in most cases the easiest solution is to stick them in the one they've been trained for.

We could also go with a third hypothesis: that *women aren't actually bizarre aliens from the planet Zygra'ax with completely inexplicable preferences*. I mean, suppose you had the following two options:

1. A job working from home, where you are your own boss. The job description is "spending as much or as little time as you want with your own children and helping them grow and adjust to the adult world." (but Sister Y also has a post on the childless alternative to this)

2. A job in the office, where you do have a boss, and she wants you to get her the Atkins report "by yesterday" or she is going to throw your sorry ass out on the street where it belongs, and there *better* not be any complaints about it this time.

Assume both jobs would give you exactly the same amount of social status and respect.

Now assume that suddenly a bunch of people come along saying that *actually*, only losers pick Job 1 and surely you're not a *loser*, are you? And you have to watch all your former Job 1 buddies go out and take Job 2 and be praised for this and your husband asks why *you* aren't going into Job 2 and contributing something to the family finances for once, and eventually you just give in and go to Job 2, but also you've got to do large portions of Job 1, and also the extra income mysteriously fails to give your family any more money and in fact you are worse off financially than before.

Is it so hard to imagine that a lot of women would be less happy under this new scenario?

Now of course (most) feminists very reasonably say that it's Totally Okay If You Want To Stay Home And We're Not Trying To Force Anyone. But let's use the feminists' own criteria on that one. Suppose Disney put out a series of movies in which they had lots of great female role models who only worked in the home and were subservient to their husbands all the time, and lauded them as *real* women who were courageous and awesome and sexy and not just poor oppressed stick-in-the-muds, and then at the end they flashed a brief message "But Of Course Working Outside The Home Is Totally Okay Also". Do you think feminists would respond "Yeah, we have no problem with this, after all they *did* flash that message at the end"?

Aside from being better for women, traditional marriages seem to have many other benefits. They allow someone to bring up the children so that they don't have to spend their childhood in front of the television being socialized by reruns of *Drug-*

Using Hypersexual Gangsters With Machine Guns. They ensure that at least one member of each couple has time to be doing things that every household should be doing anyway, like keeping careful track of finances, attending parent-teacher conferences, and keeping in touch with family.

So do men need to force women to stay barefoot and in the kitchen all the time, and chase Marie Curie out of physics class so she can go home and bake for her husband?

By this point you may be noticing a trend. No, we don't need to do that. If we stopped optimizing the media to send feminist messages as loud as possible, if we stopped actively opposing any even slightly positive portrayal of a housewife as "sexist" and "behind the times", and if we stopped having entire huge lobby groups supported vehemently by millions of people *dedicated entirely to making the problem worse*, then maybe things would take care of themselves.

There's some sort of metaphor here…something about dirt… or a shovel…nah, never mind.

Plays Well In Groups

Suppose you were kidnapped by terrorists, and you needed someone to organize a rescue. Would you prefer the task be delegated to the Unitarians, or the Mormons?

This question isn't about whether you think an *individual* Unitarian or Mormon would make a better person to rush in Rambo-style and get you out of there. It's about whether you would prefer the Unitarian Church or the Mormon Church to coordinate your rescue.

I would go with the Mormons. The Mormons seem *effective* in all sorts of ways. They're effective evangelists. They're effect fundraisers. They're effective at keeping the average believer

following their commandments. They would figure out a plan, implement it, and come in guns-blazing.

The Unitarians would be a disaster. First someone would interrupt the discussion to ask whether it's fair to use the word "terrorists", or whether we should use the less judgmental "militant". Several people would note that until investigating the situation more clearly, they can't even be sure the terrorists aren't in the right in this case. In fact, what *is* "right" anyway? An attempt to shut down this discussion to focus more on the object-level problem would be met with cries of "censorship!".

If anyone did come up with a plan, a hundred different pedants would try to display their intelligence by nitpicking meaningless details. Eventually some people would say that it's an outrage that no one's even *considering* whether the bullets being used are recyclable, and decide to split off and mount their *own*, ecologically-friendly rescue attempt. In the end, four different schismatic rescue attempts would run into each other, mistake each other for the enemy, and annhilate themselves while the actual terrorists never even hear about it.

(if it were Reform Jews, the story would be broadly similar, but with *twenty* different rescue attempts, and I say this fondly, as someone who attended a liberal synagogue for ten years)

One relevant difference between Mormons and Unitarians seems to be a cultural one. It's not quite that the Mormons value conformity and the Unitarians value indivduality – that's not exactly *wrong*, but it's letting progressives bend language to their will, the same way as calling the two sides of the abortion debate "pro-freedom" and "anti-woman" or whatever they do nowadays. It's more like a Mormon norm that the proper goal of a discussion is agreement, and a Unitarian norm that the proper goal of a discussion is disagreement.

There's a saying I've heard in a lot of groups, which is something along the lines of "diversity is what unites us". This is nice and memorable, but there are other groups where *unity* is what unites them, and they seem to be more, well, united.

Unity doesn't just arise by a sudden and peculiar blessing of the angel Moroni. It's the sort of thing you can create. Holidays and festivals and weird rituals create unity. If everyone jumps up and down three times on the summer solstice, then yes, objectively this is dumb, but you feel a little more bonded with the other people who do it: *I'm* one of the solstice-jumpers, and *you're* one of the solstice-jumpers, and that makes us solstice-jumpers together.

Robert Putnam famously found that the greater the diversity in a community:

> …the less people vote, the less they volunteer, and the less they give to charity and work on community projects. In the most diverse communities, neighbors trust one another about half as much as they do in the most homogenous settings. The study, the largest ever on civic engagement in America, found that virtually all measures of civic health are lower in more diverse settings. "The extent of the effect is shocking," says Scott Paige, a University of Michigan political scientist.

I don't think this effect is particularly related to race. I bet that if you throw together a community of white, black, Asian, Hispanic, and Martian Mormons, they act as a "non-diverse" community. As we saw before, culture trumps race.

So this sort of cultural unity is exactly the sort of thing we need to improve civic life and prevent racism…and of course, it's exactly what progressives get enraged if we try to produce.

In America, progressivism focuses on pointing out how terrible American culture is and how much other people's cultures are better than ours. If we celebrate Columbus Day, we have to spend the whole time hearing about what a jerk Columbus was (disclaimer: to be fair, Columbus was a *huge* jerk). If we celebrate Washington's birthday, we have to spend the whole time hearing about how awful it was that Washington owned slaves. Goodness help us if someone tries to celebrate Christmas – there are now areas where if a city puts up Christmas decorations, it has to give equal space to atheist groups to put up displays about how Christmas is stupid and people who celebrate it suck. That's…probably not the way to maximize cultural unity, exactly?

We are a culture engaged in the continuing project of subverting itself. Our heroes have been toppled, our rituals mocked, and one gains status by figuring out new and better ways to show how the things that should unite us are actually stupid and oppressive. Even the conservatives who wear American flag lapel pins and stuff spend most of their time talking about how they hate America today and the American government and everything else associated with America except for those stupid flag pins of theirs.

Compare this to olden cultures. If someone in Victorian Britain says "God save the Queen!", then everyone else repeated "God save the Queen!", and more important, *they mean it*. "England expects every man to do their duty" is actually perceived as a *compelling reason* why one's duty should be done.

It would seem that the Victorian British are more on the Mormon side and modern Americans more like the Unitarians. And in fact, the Victorians managed to colonize half the planet while America can't even get the Afghans to stop shooting

each other. While one may not agree with Victorian Britain's aims, one has to wonder what would happen if that kind of will, energy, and unity of purpose were directed towards a worthier goal (I wonder this about the Mormon Church too).

Reactionaries would go further and explore this idea in a depth I don't have time for, besides to say that they believe many historical cultures were carefully optimized and time-tested for unifying potential, and that they really sunk deep into the bones of the populace until failing to identify with them would have been unthinkable. The three cultures they most often cite as virtuous examples here are Imperial China, medieval Catholicism, and Victorian Britain; although it would be foolish to try to re-establish one of those exactly in a population not thoroughly steeped in them, we could at least try to make our own culture a little more like they were.

Once again, the Reactionary claim is not necessarily that we have to brainwash people or drag the Jews kicking and screaming to Christmas parties. It's just that maybe we should stop deliberately optimizing society for as little unity and shared culture as humanly possible.

Reach For The Tsars

I have noticed a tendency of mine to reply to arguments with "Well yeah, that would work for the X Czar, but there's no such thing."

For example, take the problems with the scientific community, which my friends in Berkeley often discuss. There's lots of publication bias, statistics are done in a confusing and misleading way out of sheer inertia, and replications often happen very late or not at all. And sometimes someone will say something like "I can't believe people are too dumb to fix Science. All we would have to do is require early registration

of studies to avoid publication bias, turn this new and powerful statistical technique into the new standard, and accord higher status to scientists who do replication experiments. It would be really simple and it would vastly increase scientific progress. I must just be smarter than all existing scientists, since I'm able to think of this and they aren't."

And I answer "Well, yeah, that would work for the Science Czar. He could just make a Science Decree that everyone has to use the right statistics, and make another Science Decree that everyone must accord replications higher status. And since we all follow the Science Czar's Science Decrees, it would all work perfectly!"

Why exactly am I being so sarcastic? Because things that work from a czar's-eye view don't work from within the system. No *individual* scientist has an incentive to unilaterally switch to the new statistical technique for her *own* research, since it would make her research less likely to produce earth-shattering results and since it would just confuse all the other scientists. They just have an incentive to want *everybody else* to do it, at which point they would follow along.

Likewise, no journal has the incentive to unilaterally demand early registration, since that just means everyone who forgot to early register their studies would switch to their competitors' journals.

And since the system is *only* made of individual scientists and individual journals, no one is ever going to switch and science will stay exactly as it is.

I use this "czar" terminology a lot. Like when people talk about reforming the education system, I point out that right now students' incentive is to go to the most prestigious college they can get into so employers will hire them, employers'

incentive is to get students from the most prestigious college they can so that they can defend their decision to their boss if it goes wrong, and colleges' incentive is to do whatever it takes to get more prestige, as measured in *US News and World Report* rankings. Does this lead to huge waste and poor education? Yes. Could an Education Czar notice this and make some Education Decrees that lead to a vastly more efficient system? Easily! But since there's no Education Czar everybody is just going to follow their own incentives, which have nothing to do with education or efficiency.

There is an extraordinarily useful pattern of refactored agency in which you view humans as basically actors playing roles determined by their incentives. Anyone who strays even slightly from their role is outcompeted and replaced by an understudy who will do better. That means the final state of a system is determined entirely by its initial state and the dance of incentives inside of it.

If a system has perverse incentives, it's not going to magically fix itself; no one inside the system has an incentive to do that. The end user of the system – the student or consumer – is already part of the incentive flow, so they're not going to be helpful. The only hope is that the system can get a Czar – an Unincentivized Incentivizer, someone who controls the entire system while standing outside of it.

I alluded to this a lot in my (warning: political piece even longer than this one) Non-Libertarian FAQ. I argued that because systems can't always self-improve from the inside, every so often you need a government to coordinate things.

Reactionaries would go further and say that a standard liberal democratic government is not an Unincentivized Incentivizer. Government officials are beholden to the electorate and to

their campaign donors, and they need to worry about being outcompeted by the other party. They, too, are slaves to their incentives. The obvious solution to corporate welfare is “end corporate welfare”. A three year old could think of it. But anyone who tried would get outcompeted by powerful corporate interests backing the campaigns of their opponents, or outcompeted by other states that still have corporate welfare and use it to send businesses and jobs their way. It’s obvious from outside the system, and completely impossible from the inside. It would appear we need some kind of a Government Czar.

You know who had a Government Czar? Imperial Russia. For short, they just called him “Czar”.

Everyone realizes our current model of government is screwed up and corrupt. We keep electing fresh new Washington Outsiders who promise with bright eyes to unupscrew and decorruptify it. And then they keep being exactly as screwed up and corrupt as the last group, because if you hire a new actor to play the same role, the lines are still going to come out exactly the same. Want reform? The lines to “Act V: An Attempt To Reform The System” are already written and have been delivered dozens of times already. How is changing the actors and actresses going to help?

A Czar could actually get stuff done. Imperial Decree 1: End all corporate welfare. Imperial Decree 2: Close all tax loopholes. Imperial Decree 3: Health care system that doesn’t suck. You get the idea.

Would the Czar be corrupt and greedy and tyrannical? Yes, probably. Let’s say he decided to use our tax money to build himself a mansion ten times bigger than the Palace of Versailles. The Internet suggests that building Versailles today

would cost somewhere between $200M and $1B, so let's dectuple the high range of that estimate and say the Czar built himself a $10 billion dollar palace. And he wants it plated in solid gold, so that's another $10 billion. Fine. Corporate welfare is $200B per year. If the Czar were to tell us "I am going to take your tax money and spend it on a giant palace ten times the size of Versailles covered in solid gold", the proper response would be "Great, but what are we going to do with the other $180 billion dollars you're saving us?"

(here I am being facetious. A better answer might be to point out that the British royal family already lives in a giant palace, and they by all accounts earn the country more than they cost)

As for the tyranny, we have Fnargl's shining example to inspire us. But really. Suppose Obama were named Czar. Do we really think he'd start sending Republicans to penal camps in Alaska for disagreeing with him? If Sasha took over as Czarina, do you think *she'd* do that?

Is this the face of someone who would crush you with an iron fist?

In the democratic system, the incentive is always for the country to become more progressive, because progressivism is the appeal to the lowest common denominator. There may be reversals, false starts, and Reagan Revolutions, but over the course of centuries democracy means inevitable creeping progress. As Mencius Moldbug says, "Cthulhu swims slowly, but he always swims left." A Czar, free from these incentives, would be able to take the best of progressivism and leave the rest behind.

(the Reactionaries I beta-tested this essay with say that the last paragraph deserves much more space, that there are many complicated theories of why this holds true, and that it is a central feature of Reactionary thought. I don't understand this well enough to write about it yet, but you may want to read Moldbug on…no, on second thought, just let it pass.)

So who gets to be Czar? Probably the most important factor is a Schelling point: it should be someone everyone agrees has the unquestioned right to rule. Obama is not a *bad* choice, but one worries he may be a little too progressive to treat the job with the seriousness it deserves. We could import the British monarchy, but really ever since the Glorious Revolution they've been a bit too constitutional for our purposes. If we wanted a genuine, legitimate British monarch of the old royal line, someone with authority flowing through his very veins, our best choice is, indeed to exhume the body of King James II (ruled 1685 – 1688), clone him, and place the clone on the throne of the new United States Of The Western World.

Really, it's just common sense.

A Brief Survey Of Not Directly Political Reactionary Philosophy

We have reached the goal we set for ourselves. Is this a comprehensive understanding of Reactionary thought?

No. This focuses on political philosophy, but Reaction is a complete philosophical movement with many other branches.

For example, Reactionary moral theories tend to focus on the dichotomy between Virtue and Decadence. Extensional definitions might do best here: consider the difference in outlook between Seneca the Stoic and the Roman Emperor Nero, or between Liu Bei and Cao Cao, or between Thomas More and Henry VIII. In each of these cases, a virtuous figure recognized the decadence of his society and willfully refused to succumb to it. Of course, an even more virtuous example would be someone like Lycurgus, who realized the decadence of his society and so *went out and fixed society*.

Reactionary aesthetic theories tend to be, well, reactions against progressive aesthetic theories. To Reactionaries, the epitome of the progressive aesthetic theory against which they rebel is the fairy tale of the Ugly Duckling, where one duckling is uglier than the rest, everyone mocks him, but then he turns out to be the most beautiful of all. The moral of the story is that ugly things are really the most beautiful, beautiful things are for bullies who just want to oppress the less beautiful things, and if you don't realize this, you're dumb and have no taste.

Therefore, decent, *sophisticated* people must scoff at anything outwardly beautiful and say that it's probably oppressive in some way, while gushing over anything apparently ugly. Cathedrals are "gaudy" or "tacky", but Brutalist concrete blocks are "revolutionary" and "groundbreaking". An especially conventionally attractive woman is probably just "self-objectifying" and "pandering", but someone with ten

tattoos and a shaved head is “truly confident in her femininity”. Art of the sort people have been proven to like most is old-fashioned and conformist; *real* art is urinals that artistically convey an anti-art message, or paintings so baffling that no one can tell if they are accidentally hung upside-down.

The Reactionary aesthetic, then, is something so simple that if it weren’t specifically a reaction to something that already exists, it would sound stupid: no, beautiful things are legitimately beautiful, ugly things are legitimately ugly, any attempt to disguise this raises suspicions of ulterior motives.

Reactionaries also seem to be really into metaphysics, especially of the scholastic variety, but I have yet to be able to understand this. Blatant racism, attempts to clone long-dead monarchs, and giving a gold-obsessed alien absolute power all seem like they could sort of make sense in the right light, but why anyone would want more metaphysics is honestly completely beyond me.

But Seriously, What Do We Do About This Hole? And How Fast Should We Be Digging, Anyway?

We started with an argument that modern culture probably doesn’t give us a very impartial view on the relative merits of modern culture, and so we should investigate this more thorougly.

We noted that on many of the criteria we care about, the present is better only because of its improved technology. We further noted that on other criteria, even *despite* our better technology, past societies seemed to outperform us

Nevertheless, we identified some areas where the present really did seem better than the past. The present was less racist, less sexist, less colonialist, more humane, and less jingoistic.

We then went through each of those things and showed why they might not be as purely beneficial are as generally believed. We found evidence that societies many would call "racist" give minorities better measurable outcomes; that societies many would call "sexist" give women higher self-reported life satisfaction; that colonialism led to peace and economic growth that decolonialism was unable to match; and that supposedly more "humane" policies end up torturing their victims far more than just getting something superficially cruel over wit would; and even that cultural unity, which some might call "jingoism", has been empirically shown to be an important factor in building communities and inspiring prosocial sentiment.

Therefore, we found that all the points we had previously noted as advantages of present over past societies were, when examined more closely, in fact points in the past societies' favor.

Next, we looked at how we might replicate these advantages of past societies in a world which seems to be moving inexorably further toward so-called progressive ideals. We independently came up with the same solution that these past societies used: the idea of a monarch, either constitutional or (preferably) absolutist. We found that many of the problems we would expect such a monarch to produce are exaggerated or unlikely.

Finally, we identified this ideal monarch as a clone of James II of the United Kingdom.

We also went into a survey of a couple of other Reactionary ideas. Other such ideas I have *not* included simply because I was totally unable to understand or sympathize with them and so couldn't give them fair treatment include: an obsession with

chastity, highly positive feelings about Catholicism that never go as far as actually going to church or believing any Catholic doctrine in a non-ironic way, neo-formalism, and what the heck the Whigs have to do with anything.

Nevertheless, I hope that this has been a not-entirely futile exercise in trying to Ideological Turing Test an opposing belief. I think Reactionaries are correct that some liberal ideas have managed to make their way into an echo chamber that makes them hard to examine. And even though the Reactionaries themselves are way too rightist, I think it's good to have their ideas out there in the Hegelian sense of "and then the unexamined-conservativism touched the unexamined-liberalism and in a puff of smoke they merged to magically become the perfect political system!"

_ * _ * _ * _

Once again, expect my counterargument to this sometime in the next while. I would be interested in hearing other people's counterarguments in the meantime and am very likely to steal them. I am also likely to ignore some of them if they make arguments I already agree with and so feel no need to debate, but I would still enjoy reading them. Basically I welcome comments and discussion from all sides.

With one exception. Yes, I have included the racist parts of Reactionary philosophy above. Yes, those points need to be debated, and some of that debate may be in favor. But any comment that moves away from the sort of dry scientific racism used to prove or disprove political theorems, and toward the sort where they're just shouting ethnic slurs and attacking racial groups to make their members feel bad, *will* be deleted and the person involved probably IP-banned. I also

reserve the right to edit comments that don’t quite reach that point but are noticeably in need of rephrasing.

A Thrive/Survive Theory of the Political Spectrum

I admitted in my last post on Reaction that I devoted insufficient space to the question of why society does seem to be drifting gradually leftward. And I now realize that in order to critique the Reactionary worldview effectively we're going to have to go there.

The easiest answer would be "because we retroactively define leftism as the direction that society went". But this is not true. Communism is very leftist, but society eventually decided not to go that way. It seems fair to say that there are certain areas where society did *not* go to the left, like in the growth of free trade and the gradual lowering of tax rates, but upon realizing this we don't feel the slightest urge to redefine "low tax rates" as leftist.

So what *is* leftism? For that matter, what is rightism?

Any theory of these two ideas would have to explain at least the following data points:

1) Why do both ideologies combine seemingly unrelated political ideas? For example, why do people who want laissez-faire free trade empirically also prefer a strong military and oppose gay marriage? Why do people who want to help the environment also support feminism and dislike school vouchers?

2) Why do the two ideologies seem *broadly* stable across different times and cultures, such that it's relatively easy to point out the Tories as further right than the Whigs, or ancient Athens as further left than ancient Sparta? For that matter, why do they seem to correspond to certain neural patterns in the

brain, such that neurologists can determine your political beliefs with 83% accuracy by examining brain structure alone?

3) Why do these basically political ideas correlate so well with moral, aesthetic, and religious preferences?

4) The original question: how come, given enough time and left to itself, leftism seems to usually win out over rightism, pushing the Overton window a bit forward until there's a new leftism and rightism?

I have a hypothesis that explains *most* of this, but first let me go through some proposed alternatives.

The Reactionaries have at least two theories. Moldbug suggests that rightism is common sense, and leftism is Christianity minus the religious trappings and rightism is rational thought. Another of his posts suggests that leftism is naked power-grabbing and rightism is virtuous pro-social behavior.

But the first of these fails to explain point 1; how come most traditionally Christian ideas end up on the right side of the aisle? It fails to explain 2 – how come we can call Sparta rightist even in the pre-Christian age? It might explain 3. But it definitely fails point 4; even if it were true, why would this weird neo-Christian sect suddenly take off just as all other Christian sects are hemorrhaging believers? As for the second, it explains point 4 and point 4 only, and seems, well, maybe a little completely obviously self-serving?

The Libertarians say that leftism supports government intervention on economic but not social issues, and rightism supports government intervention on social but not economic issues. Unfortunately, this isn't really true. Leftists support government intervention in society in the form of gun control, hate speech laws, funding for the arts, and sex ed in schools. In

fact, leftists are sometimes even accused of being in favor of "social engineering". Meanwhile, conservatives lead things like the home schooling and school choice movements, which seem to be about *less* government regulation of society. Having gotten Point 1 not quite right, this theory then goes on to completely ignore points 2, 3, and 4.

The scientists studying neuropolitics in that article I linked to say things like "Liberals tend to seek out novelty and uncertainty, while conservatives exhibit strong changes in attitude to threatening situations. The former are more willing to accept risk, while the latter tends to have more intense physical reactions to threatening stimuli." But this seems flawed. Leftists have an intense physical reaction to the threatening situation of global warming. Rightists seek out the novelty and accept the risk of a foreign war that *might* increase America's global power at minimal cost but *might* waste hundreds of thousands of lives to no end. Another failure of 1, I'll give it 2 or 3, and once again no love for point 4.

Okay, I'll put you out of your misery and tell you my hypothesis now. My hypothesis is that rightism is what happens when you're optimizing for surviving an unsafe environment, leftism is what happens when you're optimized for thriving in a safe environment.

The Dead Have Risen, And They're Voting Republican

Before I explain, a story. Last night at a dinner party we discussed Dungeons and Dragons orientations. One guest declared that he thought Lawful Good was a contradiction in terms, very nearly at the same moment as a second guest declared that he thought Chaotic Good was a contradiction in terms. What's up?

I think the first guest was expressing a basically leftist world view. It is a fact of nature that society will always be orderly, the economy always expanding. Crime will be a vague rumor but generally under control. All that the marginal unit of extra law enforcement adds to this pleasant state is cops beating up random black people, or throwing a teenager in jail because she wanted to try marijuana.

The second guest was expressing a basically rightist world view. The prosperous, orderly society we know and love is hanging *by a frickin' thread*. At any moment, terrorists or criminals or just poor management could destroy everything. It is *really really good* that we have police in order to be the "thin blue line" between civilization and chaos, and we might sleep easier in our beds at night if that blue line were a little thicker and we had a little more buffer room.

I propose that the best way for leftists to get themselves in a rightist frame of mind is to imagine there is a zombie apocalypse tomorrow. It is a very big zombie apocalypse and it doesn't look like it's going to be one of those ones where a plucky band just has to keep themselves alive until the cavalry ride in and restore order. This is going to be one of your *long-term* zombie apocalypses. What are you going to want?

First and most important, guns. Lots and lots of guns.

Second, you're going to have a deep and abiding affection for the military and the police. You're going to hope that the government has given them a *lot* of funding over the past few years.

Third, you're going to start praying. Really hard. If someone looks like they're doing something that might offend God, you're going to very vehemently ask them to stop. However few or many atheists there may be in foxholes, there are

probably fewer when those foxholes are surrounded by zombies. Or, as Karl Marx famously said of zombie uprisings, "Who cares if it's an opiate? / It's time to pray!"

Fourth, you're going to be extremely suspicious of outsiders. It's not just that they could be infected. There are probably going to be all sorts of desperate people around, looking to steal your supplies, your guns, your ammo. You trust your friends, you trust your neighbors, and if someone who looks different than you and seems a bit shifty comes up to you, you turn them away or just kill them before they kill you.

Fifth, you're going to want *hierarchy* and *conformity*. When the leader says run, everyone runs. If someone is constantly slowing the group down, questioning the group, causing trouble, causing dissent, they're a troublemaker and they can either shut up or take their chances on their own. There's a reason all modern militaries work on a hierarchical system that tries to maximize group coherence.

Sixth, you are not going to be sentimental. If someone gets bitten by the zombies, they get shot. Doesn't matter if it's really sad, doesn't matter if it wasn't their own fault. If someone breaks the rules and steals supplies for themselves, they get punished. If someone refuses to pull their weight, they get left behind. Harsh? Yes. But there's no room for people who don't contribute in a sleek urban postapocalyptic zombie-fighting machine.

Seventh, you want to *maximize wealth*. Whatever gets you the supplies you need, you're going to do. If that means forcing people to work jobs they don't like, that's the sacrifice they've got to make. If your raid on a grocery store leaves less behind for everyone else, well, that's too bad but you need the food. Are woodland animals going to go extinct as more and more

survivors retreat to the woods and rely on them for food? That's not the kind of thing you're worried about when you're half-starved and only a few hours ahead of the zombie horde.

Eighth, strong purity/contamination ethics. We know that purity/contamination ethics are an evolutionary defense against sickness: disgusting things like urine, feces, dirt, blood, insects, and rotting corpses are all vectors of infection; creepy animals like spiders, snakes, and centipedes are all vectors for poisoning. Maybe right now you don't worry too much about this. But in a world where the hospitals are all overrun by zombies and you need to outrun a ravenous horde at a moment's notice, this becomes a much bigger deal. Not to mention that anything you catch might be the dreaded Zombie Virus.

Ninth, an emphasis on practical skills rather than book learning. That eggheaded Professor of Critical Studies? Can't use a gun, isn't studying a subject you can use to invent bigger guns, not a useful ally. Probably would just get in the way. Big masculine men who can build shelters and fight with weapons are useful. So are fertile women who can help breed the next generation of humans. Anyone else is just another mouth to feed.

Tenth, extreme black and white thinking. It's not *useful* to wonder whether or not the zombies are only fulfilling a biological drive and suffer terribly when you kill them despite not being morally in the wrong. It's *useful* to believe they're the hellish undead and it's your sacred duty to fight them by any means necessary.

In other words, "take actions that would be beneficial to survival in case of a zombie apocalypse" seems to get us rightist positions on a lot of issues. We can generalize from

zombie apocalypses to any desperate conditions in which you're not sure that you're going to make it and need to succeed at any cost.

What about the opposite? Let's imagine a future utopia of infinite technology. Robotic factories produce far more wealth than anyone could possibly need. The laws of Nature have been altered to make crime and violence physically impossible (although this technology occasionally suffers glitches). Infinitely loving nurture-bots take over any portions of child-rearing that the parents find boring. And all traumatic events can be wiped from people's minds, restoring them to a state of bliss. Even death itself has disappeared. What policies are useful for this happy state?

First of all, we probably shouldn't have a police force. Given that crime is impossible, at best they would be useless and at worst they might go around flexing their authority and causing trouble.

Second, religion seems kind of superfluous. Throughout history, richer civilizations have been less religious and our post-scarcity society should be no exception. What would you pray for? What fear is there for faith to allay? With vast supercomputers that know all things, what lingering questions are there for the Bible to answer?

Third, assuming people still have jobs or something, we should probably make them as nice as possible. It doesn't matter if it hurts productivity; we're producing far more than we need anyway. We should enforce short work hours and ample maternity and paternity leave so that everyone has time to concentrate on the more important things in life.

Fourth, interest in the environment. We have no shortage of material goods; if our lives lack anything it is beauty and

connection to nature. So it will be nice to have as many pleasant green spaces as possible; and if this means a little less oil, it's not like our Oil-Making-Machines can't make up the extra.

Fifth, free love. There's no worries about STDs, the family unit isn't necessary for any kind of economic survival, and the nurture-bots and trauma-erasure-centers can take care of the kids of anything goes wrong. And since we don't have anything else to do, we might as well enjoy ourselves with infinite sex.

I was going to go for ten here too, but you get the picture. This world of infinite abundance is a great match for leftist values. I imagine even a lot of rightists and Reactionaries would be happy enough with leftism in a situation like this.

I should also mention what would no doubt be the main pastime of the people of this latter world: signaling.

When people are no longer constrained by reality, they spend most of their energy in signaling games. This is why rich people build ever-bigger yachts and fret over the parties they throw and who got invited where. It's why heirs and heiresses so often become patrons of the art, or donors to major charities. Once you've got enough money, the next thing you need is status, and signaling is the way to get it.

So the people of this final utopia will be obsessed with looking good. They will become moralists, and try to prove themselves more virtuous than their neighbors. Their sophistication will gradually increase as each tries to establish themselves as a critic, as tasteful, as a member of an aristocracy that can no longer be defined in terms of money. They will become conniving, figuring out ways to raise their own social status at their neighbors' expense. Or they will devolve into a host of

competing subcultures, united only by their pride in their defiance of a "norm" which is quickly ceasing to exist.

Chris wrote this comment to my last post's section on Reactionary aesthetics:

> The things Reactionaries complain about in aesthetics seem not the fault of progressives, but the result of an unavoidable signaling logic. See Quentin Bell on what he called "conspicuous outrage."

I agree with Chris 100% here, but I don't think this is opposed to the Reactionaries' link between this aesthetic and leftism. I think that leftists are the sort of people who are so secure that they can start thinking about how to excel at signaling games.

An Evaluation of the Thrive/Survive Theory

This is close to an explanation of our Point 1. It does not quite explain all left vs. right positions (in particular I despair of *any* theory that will tell me why school choice is a rightist rather than a leftist issue) but it does as well as any of the others, and better than some.

This also satisfies Point 2. The distinction between security and insecurity is far older than Classical Greece; it is perfectly reasonable for Athenian society to start from the assumption of the one and the Spartans to go with the other.

I admit some confusions. For example, it seems weird that poor people, the people who are *actually* desperate and insecure, are often leftist, whereas rich people, the ones who are *actually* completely safe, are often rightist. I would have to appeal to economic self-interest here: the poor are leftist because leftism is the philosophy that says to throw lots of resources at helping the poor, and the rich are rightist because rightism says to let the rich keep getting richer. Despite voting

records, I expect the poor to share more rightist social values (eg be more religious, more racist) and the rich to to share more leftist social values (more intellectual as opposed to practical, less obsessed with guns). For a more comprehensive theory of economic self-interest and politics, see my essay on the subject.

This theory also satisfies Point 3. Developmental psychology has gradually been moving towards a paradigm where our biology actively seeks out information about our environment and then toggles between different modes based on what it finds. Probably the most talked-about example of this paradigm is the thrifty phenotype idea, devised to explain the observation that children starved in the womb will grow up to become obese. The idea is that some system notices that there seems to be very little food, and goes into "desperately conserve food" mode, which when food becomes more plentiful leads to obesity.

Another example, more clearly neurological, is the tendency of children who grow up in broken homes to have poor life outcomes. Although this was originally just interpreted as "damage", an equally valid theory is that the brain seeks out information on what kind of society it lives in – one based on love and trust, or one based on violence and mistrust – and then activates the appropriate coping strategy. If child abuse or something makes the brain conclude we live in a violence and mistrust society, it alters its neural architecture to be violent and mistrustful – and hence dooms itself to future bad outcomes.

It seems broadly plausible that there could be one of these switches for something like "social stability". If the brain finds itself in a stable environment where everything is abundant, it sort of lowers the mental threat level and concludes that

everything will always be okay and its job is to enjoy itself and win signaling games. If it finds itself in an environment of scarcity, it will raise the mental threat level and set its job to "survive at any cost".

What would toggle this switch? My guess is that genetics plays a very large role in setting the threshold (explaining why party affiliation is highly heritable) and that a lot of the remainder is implicit messages we get in childhood from our parents, school, church, et cetera. Actual rational argument and post-childhood life experiences make up the last few percent of variation.

Knowing this, the answer to Point 4 is blindingly obvious. Leftism wins over time because technology advances over time which means societies become more secure and abundant over time.

As a decent natural experiment, take the Fall of Rome. Both Greece and Rome were *relatively* leftist, with freedom of religion, democratic-republican governments, weak gender norms, minimal family values, and a high emphasis on education and abstract ideas. After the Fall of Rome, when Europe was set back technologically into a Dark Age, rightism returned with a vengeance. People became incredibly religious, militant, pragmatic, and provincial, and the government switched to an ad hoc and extremely hierarchical feudalism. This era of conservativism ended only when society reached the same level of technology and organization as the Greeks and Romans. So it's not that cultures become more leftist over *time*, it's that leftism varies with social and economic security.

Both rightists and leftists will find much to like in this idea. The rightists will ask: "So you mean that rightism is optimized

for survival and effectiveness, and leftism is optimized for hedonism and signaling games?" And I will mostly endorse this conclusion.

On the other hand, the leftists will ask: "So you mean rightism is optimized for tiny unstable bands facing a hostile wilderness, and leftism is optimized for secure, technologically advanced societies like the ones we are actually in?" And this conclusion, too, I will mostly endorse.

Given that we are in conditions that seem to favor leftist ideals, the modern debate between leftists and rightists is, to mix metaphors atrociously, about how hard we can milk the goose that lays the golden eggs. Leftists think we can just keep drawing more and more happiness and utility for all out of our massive scientific and technological progress. Rightists are holding their breath for something to go terribly, terribly wrong and require the crisis-values they have safeguarded all this time – which is why posts like this one seem to be the purest expression of rightist wish-fulfillment fantasy.

I will only remark that one of the most consistent findings of my researches through economic and political history has been the remarkable, almost supernatural resilience of our particular aureate waterfowl. To a leftist, this is good news. To a rightist, I suppose this would just be evidence of how shockingly audacious we must be to try to push our luck even *further*.

EDIT: People are taking this as pro-Reactionary. I meant it to be at least *suggestive* of anti-Reactionary ideas. See my reply to the first comment below.

We Wrestle Not With Flesh And Blood, But Against Powers And Principalities

> Mimes, in the form of God on high mutter and mumble low
> And hither and thither fly – mere puppets, who come and go
> At bidding of vast formless things shifting scenery to and fro
>
> **– an excerpt out of *Ligaea*, by Edgar Allen Poe**

There should be a post debating Reactionaries' assumptions about the superiority of past cultures and methods. Eventually I hope to write that post. But this is not it. This is the post where I claim that, *even granting all of those assumptions*, Reaction is somewhere between wrong and impossible. Why?

To borrow Poe's terminology, history as we learn it in school tends to concentrate on the puppets and ignore the vast formless things.

In a previous essay, I mentioned a pattern of refactored agency in which human beings lack agency and merely respond to incentives. I said they were "actors" reading from the "script their incentives wrote for them", and anyone who deviated from the part would be outcompeted and replaced.

This seems to broadly describe most historical figures. If Christopher Columbus had decided not to explore America, Cabral or Cabot or someone would have. Caravels existed, people needed a new trade route to India, the only question was who was going to be first.

But the puppetry expands past individuals toward whole empires and movements. If God reached into the year 1900 and removed every single Communist, and every Communist book, and erased all memory of Communism, I think it would take about five minutes before someone reinvented something much like the movement, because there were a bunch of very poor people who felt desperate and cheated crammed up against a bunch of very rich people who weren't afraid to flaunt their wealth. The new movement might have differed from Communism in minor details – maybe their color would have been blue instead of red – but it wouldn't be hard to identify.

So much for the puppets. What are these Vast Formless Things giving them their orders? I mentioned liking *Guns, Germs and Steel*, and I think Diamond has done a good job of proving geography has important historical effects. But geography is fixed, not exactly the sort of thing that's going to cause revolutions. So after that last post you probably won't be surprised to hear I think the vastest and most formless Vast Formless Thing of all is technological progress.

Engines Alone Turn The Wheels of History

The largest and furthest-reaching political changes of all time have invariably been the effect of technological progress. The largest of these, the transition from egalitarian bands to the ultrahierarchical divine monarchies of the Bronze Age, seems to have been mostly the effect of the Agricultural Revolution and its corollaries. Without committing to what order these things happened in:

– Need for a guarantee that the crops you planted will still be yours at harvest time inspires idea of private property
– Sedentary lifestyle + concept of property allow accumulation

of wealth
– Accumulation of wealth requires law enforcement to protect wealth
– Excess food allows specialization of labor
– Requirement for law enforcement + specialized labor leads to creation of warrior caste
– Powerful warrior caste + everyone else being farmers and losing the martial skills they enjoyed as hunters leads to warrior caste taking over.
– Need for large irrigation/flood control projects in many areas leads to very centralized government

And a lot of these late Neolithic/early Bronze Age cultures turned out the same way. If Ramesses II, Montezuma II, and Agamemmnon went to lunch together, they'd have a lot to talk about, despite being separated by continents and millennia. This suggests that the Generic Bronze Age Government – a god-king served by a bunch of warrior-nobles, plus massive militarism and slavery – probably just made sense given the circumstances.

I don't want to sound too deterministic and spooky here, but I do think governments have a good way of kind of converging to a local optimum. Ramesses II may not have thought "You know, the Nile floods a lot, so I should institute a strong centralized government with lots of slavery", but some people tried some things, other people tried other things, the things that worked won out, the things that didn't passed into the dustbin of history, and we got Ancient Egypt. If God reached into history and tried to turn Ancient Egypt into modern day Sweden, it wouldn't work any better than His attempt to remove Communism did a few paragraphs ago – within a few years they'd be back to worshipping Pharaohs and invading Canaanites

After the Neolithic, one of the most clear-cut examples of technology changing social structure was the fall of feudalism. Feudalism was based on a very simple calculation: one armored knight could defeat an arbitrary number of untrained peasants. To be an armored knight took your standard 10,000 hours of training; it wasn't something you can do as a side job. So once again you have at least two castes – the warrior caste and the support-the-warriors caste; since the warrior caste is both smaller and stronger, you end up with an aristocratic system. If you want to govern large territories under an aristocratic system and you don't have real-time communication, you come up with something like feudalism. And sure enough, we have pretty much the exact same social structure in medieval Europe and Sengoku Japan.

Then some new weapons were invented: pikes, longbows, crossbows, but especially firearms. Now you can get someone who *hasn't* trained 10,000 hours, give them a few days of weapon training, hand them a gun or a crossbow or something, and they can kill an armored knight. Now the power doesn't belong to the people with the best connections among the warrior nobility, it belongs to the people with enough money to hire soldiers and supply them with guns. It took a long time to realize this, especially since guns weren't that good to begin with, but when people finally got it into their heads feudalism went caput.

The printing press was an even bigger deal. I don't have my Big List O' Unbelievable Printing Press Statistics handy here, but the Internet reminds me that there were 30,000 books – total! – in Europe before the invention of the printing press. Fifty years later, 300,000 copies *just of Martin Luther's religious tracts* were printed *in a single year alone*. Among just the simpler and more direct effects:

– Protestant Reformation. Easy one. Lots of people had tried challenging the Catholic Church before, but not only could they not get their message out, but most people weren't ready for it – only the richest of the rich could even own their own Bible. Basically *as soon as* the printing press was invented this took off.
– Newspapers. All of a sudden, people who aren't the highest ranks of the nobility know what's going on at court. Some people have opinions on this. Start of modern politics where the masses know what's going on and might complain.
– The Renaissance. All these old Greek and Roman texts are spread. People realize that there are other ways to organize society beyond their own.
– Scientific Revolution. If a scientist discovers something, he can actually sent his work to other scientists in an efficient way, who can then build upon it. This was absolutely not the case for previous scientists, which is why not much happened during those periods.
– Rise of nationalism. Ability of common people to read books means more books printed in vernacular instead of Latin. This causes insular language-based communities which then feed upon themselves to become more delineated nation-states.

I was going to go into the same depth about the Industrial Revolution and the Sexual Revolution (by which I mean near-simultaneous discovery of birth control pills and antibiotics effective against syphilis), but this section is getting long, so if you promise to just agree they Changed Everything I'll make life easier for both of us and move on.

Forget King James II, Try King Canute

So the biggest changes in history have been predetermined reactions to different technological conditions. This should worry Reactionaries for several reasons.

First, I previously claimed that if Communism disappeared it would be immediately reinvented. If Ancient Egypt had randomly switched to modern Sweden, the realities of life in the Nile flood plain and of Bronze Age technology would have caused it to switch back without even breaking its stride.

I think my claim here is that cultures and ideologies have a sort of homeostatic regulatory mechanism that fits them to their conditions. This is why all Bronze Age cultures converged upon divine monarchies, and all medieval empires converged upon feudalism, and proooobably why all modern cultures converge upon liberal democracy.

Countries that avoid liberal democracy usually regret it. China would be a good example. They tried being really Communist for a while and ended up becoming an economic basketcase. If they wanted to compete on the international stage they realized they needed a stronger economy, and so liberalized their market. A competitive market requires information access, so the Chinese got access to lots of foreign media; I recently learned that any business that wants to pay for it can even legally avoid the Great Firewall. The Internet meant the Chinese could coordinate protests on microblogging platforms, leading to a bunch of riots, leading to an attempt to liberalize the system and crack down on corruption which is still going on. I'm not going to claim that China is definitely going to end up as a democracy, but I think whatever it does end up as is going to be a whole lot more like 2013 USA than like 1963 China.

China didn't *plan* to approach the Western model of government. It was just what happened to them automatically when they wanted their country to stop being a hellhole. The same is happening now in Burma, somewhat more slowly in Cuba, and in other places around the world. Even the countries

skipping the “democracy” part have been aping the Industrial Revolution, womens’ rights, and so on.

This is probably because many features of liberal democracy are adaptations to our current technological climate. For example, women’s lib seems like an adaptation both to the Sexual Revolution and to the demographic transition where people are no longer having like twenty children all the time. Representative government seems like an adaptation to mass media that allows everyone to be aware of, and usually upset about, what the country’s leadership is doing.

If you like these things, you can call it cultural evolution and assume we’re approaching some great goal of perfection. If you don’t like them, you can call them patches, such that once the demographic transition screws up traditional gender roles, we need women’s lib as a patch to contain the damage. Either way, you better not take off that patch.

So this is my first beef with Reactionaries. They see someone identifying as Progressive saying something – Gloria Steinem pushing for women’s rights or something – and they say “Oh no, that awful Progressive Gloria Steinem is screwing up our traditional gender roles. If only she would be quiet, everything would go back to normal!”

Gloria Steinem is a puppet. If she’s part of some movement, even a large saecular movement calling itself Progressivism, they, too, are puppets. It is stupid to get upset at puppets. If you rip them up, the puppeteer will get new ones.

If you don’t like women’s lib, your enemy isn’t Gloria Steinem. Your enemy is the Vast Formless Thing controlling Gloria Steinem. In this case, that would be the demographic transition.

You might be able to beat Gloria Steinem in a fight, but you can't beat the demographic transition. Or if you can, it's going to be through something a lot more complicated than going on a soapbox and condemning it, more complicated even than becoming Czar and trying to pass laws to reverse it.

King Canute tried to order back the tide. It was a dumb idea, but in his defense, it was basically just a religious spectacle so he could wax poetic about the power of God. What's *your* excuse?

Amid These Dark Satanic Mills

In the comments to the Enormous Planet-Sized Nutshell post, some people did a good – though not unassailable – job of picking apart some common Reactionary arguments for superior outcomes among past cultures. The crime data may be an artifact, and more believable homicide data suggests the modern era is safer. Modern students may learn different things than are tested on that Harvard exam which are equally valuable.

Whatever. Let's assume the Reactionaries are totally right. Past was a thousand times better than the present in every way. So what?

The past contained things like "everyone living in close-knit mono-ethnic villages". We could, perhaps, with great effort and not a little atrocity, restore the "mono-ethnic". But the close-knit? The villages? Unless we're going to roll back the Industrial Revolution, the *main* ingredient of that particular transition, the move to urbanization, is there to stay.

Any statistic in which the present differs from the past is much more likely to be a result of technology than of politics. Reactionaries correctly use this to excuse themselves of

advantages like the present's better health care or greater wealth.

But they have to acknowledge that the same manuever relieves the other side of a lot of their burdens as well. Progressives also have some uncomfortable statistics, usually those relating to social cohesion and trust and happiness. And *I am totally willing to throw every one of these out*. Of course the move to an urban society is going to do that! Of course having people work factory or office jobs instead of either on the land or in an skilled trade like blacksmithing is going to alienate them. Of course having the average person watch TV four hours a day because it's a novel superstimulus is going to affect community ties!

I suspect that the most valuable features of past societies – the ones that we read fantasy books to recapture, the ones that make Renaissance Faires and Medieval Times so attractive – have nothing to do with politics and cannot be restored through politics. In order to regain them, you're going to have to roll back the Industrial Revolution. Needless to say, that makes fighting against the demographic transition look easy.

Perfectly Prepared For A Situation That No Longer Exists

The third and last and most important point I want to bring up involves well-adaptedness.

I often hear Reactionaries make an argument like: the old ways are the result of thousands of years of trial-and-error. Those thousands of years created a remarkably stable culture that survived for centuries. When Progressives throw them out, they are abandoning something we know works for some sort of grand experiment that might end in complete failure.

And I wonder: have these people ever updated a computer program before?

I mean, take Windows 3.11. We know all about Windows 3.11. People had a long time to test it, discover its bugs, find its security holes. Windows 8, on the other hand, is totally new. Goodness only knows what sort of unpleasant surprises are lurking there.

But imagine I decided to uninstall Windows 8 from my computer and replace it with Windows 3.11. Most of my programs aren't written for Windows 3.11 and they wouldn't work. Windows 3.11 probably has no idea what to do with Wi-Fi. It probably can't handle the dual cores of my laptop. Most likely it would ask me to insert floppy disks during the installation and my computer doesn't have a floppy disk drive.

Even if Windows 3.11, with 1992 programs, on a 1992 machine, is more stable than Windows 8, with 2013 programs, on a 2013 machine – even so, Windows 3.11 with 2013 programs on a 2013 machine would be a total disaster.

I tend to agree with Reactionaries that cultures have a mechanism that gradually adapts them to their conditions. This may not be morally good – if the conditions are "cotton is very lucrative" then the "evolutionarily advantageous" adaptation for a society may be to institute slavery – but they are at least effective and stable.

But a 1600s culture with 2013 technology would be like Windows 3.11 on a 2013 computer: a complete mismatch and a complete disaster. No matter how well Bourbon France was adapted to the 1600s, it would have *no idea* what to do with 2013. If it tried, it would probably end up converging towards the same 2013-technology equilibrium – liberal democracy – as everyone else in 2013. Maybe Louis XIV could stick around as a figurehead or something.

The Reactionaries are correct that we live in a scary time, a time when changes in technology are way outpacing our ability to have any idea how to cope as a society. Maybe if you froze technology at 2013 levels for a hundred years, we would get a pretty good idea what to do with it and would build a culture as well-adapted to our technology level as the Bourbon French were to theirs.

But, uh, getting rid of our culture and replacing it with Bourbon France doesn't shortcut that process. We have a four hundred year head start over Bourbon France in adapting to *our* conditions. If we suddenly became Louis XIV, we'd just be even further behind the adaptation curve, having to reach liberal democracy first before we could get to wherever we're going.

I don't think Bourbon France was more successful, as a society, than our society is. But if you convinced me otherwise, it wouldn't make a shred of difference. Bourbon France + modern tech levels is a society that has never existed and which, I suspect, would be about as successful as Windows 3.11 trying to run Minecraft.

But Seriously, Why Did This Gaping Crack In The Earth Just Open Up? And Why Are You Yelling At The Kid With A Plastic Shovel Next To It?

Our goal was to show that, even granting Reactionaries all their assumptions about the superiority of past civilizations, trying to restore them is impossible.

We noted that the driving force of large-scale historical change was technological progress. That societies underwent cultural evolution into forms that were most adapted to the technological conditions of their age. That this evolution was convergent, and even unconnected civilizations like Ramesses'

Egypt and Montezuma's Aztecs could come to resemble each other when they faced similar material problems.

Then we noted that what looks like political progress from the outside is just humans reacting to the shifting landscape of incentives. Although feminism appears as a movement spearheaded by particular feminists who got it into their head thats feminism was a good idea and so decided to push it, a causally useful etiology of feminism would trace the technological conditions that predestined it to arise and succeed.

We accused Reactionaries of condemning or excusing such movements as if they were contingent human creations, and of acting like pushing a few humans or institutions out of the way here or there would change them. Instead, we concluded that they were vast tides in the affairs of (wo)men, and that any attempt to order them around was hubris worthy of King Canute.

Then we accused Reactionaries of a bit of a double-standard, excusing traditional societies' lesser wealth and health by placing the blame on technological progress, but being unwilling to let Progressives do the same in areas where technological progress has inevitably made us worse off, such as the production of feelings of social alienation.

Finally, we accused Reactionaries of arguing that past societies were well-adapted, without specifying well-adapted to *what*. We hypothesized that if forced to finish this statement, it would end up with "well-adapted to the technologies and conditions of the centuries they flourished". The very fact that they stopped flourishing and were replaced by our society suggest they are less well-adapted to conditions today. Or, as G.K. Chesterton puts it in a different context:

> There is one broad fact about the relations of Christianity and Paganism which is so simple that many will smile at it, but which is so important that all moderns forget it. The primary fact about Christianity and Paganism is that one came after the other. Mr. Lowes Dickinson speaks of them as if they were parallel ideals–even speaks as if Paganism were the newer of the two, and the more fitted for a new age. He suggests that the Pagan ideal will be the ultimate good of man; but if that is so, we must at least ask with more curiosity than he allows for, why it was that man actually found his ultimate good on earth under the stars, and threw it away again.

I do not think these problems completely disprove Reaction. They merely wall off several potential lines of argument in its support: the argument that ancient cultures empirically achieved better outcomes than our own, and the argument that they were more stable and better adapted.

To save Reaction, you would have to try one of the following paths.

First, you could claim that there's no such thing as cultural evolution, that cultures don't gradually become more adapted to their conditions via time. This seems plausible, but then the Reactionaries lose their own strongest argument; that older cultures were better adapted. Nevertheless, this is where I think a lot of the remaining probability of Reaction being true would be, and many of the arguments in my pro-Reaction post before continue to stand in this case.

Second, you could agree that cultures evolve, but that for some reason the cultural evolution mechanism has gone berserk over the past few hundred years. To make this stick, you'd have to give some reason this would happen. Then you'd have to

prove that it was *so berserk* that the best we could do is reboot from a saved copy from before its breakdown, even knowing that this will be completely unsuited for modern life.

Third, you could posit that for some reason cultural evolution previously drove us in a Progressive direction, but now it is driving us back in a Reactionary direction, and that you are a legitimate priest of the Vast Formless Things just making their new and revised will known unto man. To make this work, you'd have to figure out exactly when and why the Vast Formless Things changed their minds.

For most of the rest of this sequence I'll be concentrating on option 1, unless a horde of Reactionaries appear in the comments and tell me they have totally considered this problem before and 2 or 3 is the more commonly accepted view. In option 1, by sort of a coincidence past societies happened to be better than ours, and for coincidental reasons ours went off track. The onus then would be to determine which of our society's policies are or aren't bad, and what was the last stable copy of them to reboot from.

Poor Folks Do Smile… For Now

I got the opportunity to talk to GMU professor and futurist Robin Hanson today, which I immediately seized upon to iron out the few disagreements I still have with someone so brilliant. The most pressing of these is his four year old post Poor Folks Do Smile, in which he envisions a grim Malthusian future of slavery and privation for humanity and then soundly endorses it. As he puts it:

> Our robot descendants might actually be forced to slave near day and night, not to feed kids but just to pay for their body rent, their feed-stocks, their net connection, etc. Even so they'd be mostly happy.

Robin seems to be a total utilitarian who has no objections to the Repugnant Conclusion. It's a consistent position on its own (though distasteful to me), and taken at face value it has something to recommend it. I've been to some horrible places like Haiti. The Haitians have it tough, but they still sing and dance, they still love each other, they still have hopes and dreams. If the far future is Haiti with better sanitation, it wouldn't necessarily be the worst thing in the world.

But Robin has a slightly higher bar here. He believes that the near future promises advances in the uploading of human minds to computers, creating cyber-organisms he calls "ems" for "emulated humans". Ems will have many advantages over biologicals – less need for space and resources, possible elimination of biological need for sleep, and can be copied-pasted at will. A future of zillions of Malthusian ems competing for hardware and computing power is a little

different from zillions of biological humans competing for land and food.

So here is my dialogue with Robin as I remember it. I didn't take notes, so it's probably a bit off, and I'm rewriting me being confused and ummming and errrrring and meandering for a while as me having perfectly flowing rational arguments with carefully considered examples. I think I understood Robin well enough to be able to put down what he said without accidentally strawmanning him but I do notice he was much more convincing and I was much more confused and challenged in person than it looks in this transcript, so perhaps I failed. Nevertheless:

Scott: In "Poor Folks Do Smile", you say that a future of intense competition and bare subsistence will be okay because we will still have the sorts of things that make life valuable. But in a future of ems, won't there be competitive advantage to removing the things that make life valuable?

Robin: What do you mean?

Scott: Suppose you have some ems that are capable of falling in love and some that aren't. The ones that fall in love spend some time swooning or writing poetry or talking to their lover or whatever, and the ones that don't can keep working 24-7. Doesn't that give the ones that can't fall in love enough of a competitive advantage that the ones that can will be outcompeted and destroyed and eventually we'll end up with only beings incapable of love?

Robin: You can't just remove love from a human brain like that. There's no one love module.

Scott: It's probably very hard to remove love from a human brain without touching anything else. But given that the future is effectively infinitely long, and that in a world of perfect

competition it would be advantageous to do this, surely someone will succeed eventually.

Robin: Yes, the future is infinitely long. But you're speculating post-Singularity here, and the whole point of the Singularity is that it's impossible to speculate on what will happen after it. I speculate on the near and medium term future, but trying to predict the very long-term future isn't worth it.

Scott: I agree we can't predict the far future, but this is less a prediction than an anti-prediction. An anti-prediction is…wait, am I doing that thing where I explain something you invented to you?

Robin: No, I didn't invent anti-predictions. Go on.

Scott: An anti-prediction is…gah, I wish I could remember the canonical example…an anti-prediction is when you just avoid privileging a hypothesis and this sounds like a bold prediction. For example, suppose I predict with 99%+ confidence that the first alien species we meet will not be practicing Christians. In a certain context, this might sound overconfident – aliens could be atheists or Christians or Muslims, we don't really know, but since I don't know anything at all about aliens it sounds overconfident to be so sure it won't be the Christian one. But in fact this is justified, since Christianity is just a tiny section of possible-religion-space that only seems important to us because we know about it. The aliens' likelihood of being Christian isn't 1/3 ("either Christian, or atheist, or Muslim) but more like 1/1 trillion (Christianity out of the space of all conceivably possible religions). The only way the aliens could be Christian is if it was for some reason correlated with our own civilization's Christianity, like we went over there to convert them, or if

Christianity was true and both us and the aliens were truth-seekers. My point is that human values, like love, are a tiny fraction of mindspace. So saying that the far future won't have them is an antiprediction.

Robin: Values like love were selected by evolution. We can expect that similar selection pressures in the future will produce, if not the same values, ones that are similar enough to be recognizable.

Scott: The hypercompetitive marketplace of an advanced cybernetic civilization is different enough from an African savannah that I really don't think that's true. Love evolved in order to convince people to reproduce and raise children. If ems can reproduce by copy-pasting and end out with full adults, that's not a society that will replicate the need for love.

Robin: Love is useful for a lot of other things. Probably the same mental circuitry that causes people to fall in love is the sort of thing you need to make people love their work and stay motivated.

Scott: Antiprediction! Most mind designs that can effectively perform tasks don't need circuitry that also causes falling in love!

Robin: The trouble with this whole antiprediction concept is…so what if I told you that in the far future, people would travel much faster than light. Would that be an antiprediction? After all, most physical theories don't include a hard light-speed limit.

Scott: The trouble with traveling faster than light is that it's physically impossible. Are you trying to make the claim that a mind design that doesn't include something like human love is physically impossible?

Robin: I'm trying to make the claim that it's not something you can plausibly get to by modifying humans.

Scott: Fine. Forget modifying humans. People just try to build something new and more efficient from the ground up.

Robin: Maybe in the ridiculously far future…

Scott: But we both agree on a sort of singularitarian world-view where "history is speeding up". The "ridiculously far future" could be twenty years from now if ten years from now they invent ems that can be run at a hundred times normal speed. If the ridiculously far future aka twenty years from now is one where human values like love are completely absent, that seems…really bad. And if we want to prevent it, it seems like that goes through trying to prevent a "merely" Malthusian medium-term future in which people are effective slaves but we haven't *quite* figured out how to hack out love yet.

Robin: Attempting to influence the far future is very dangerous. In most cases we can't predict the long-term consequences of our action. The near future will be in a much better position to influence the far future than we are. My claim, which you don't seem to disagree with, is that the near future will be non-hellish and preserve human values like love. Let's let this near future figure out whether the far future will be unacceptable. As time goes on, people gain better ability to coordinate, so the near future should be better at fixing our problems anyway.

Scott: As time goes on people gain better ability to coordinate?

Robin: Yes. In the old days, most decisions were made at the village or provincial level. Now we're gradually centralizing decisions to the national and often even the supranational

level. The modern world is much more effective at coordinating solutions to its problems than the past.

Scott glances at Michael Anissimov, probably the most vocal Reactionary in Berkeley, who has been standing there listening to the conversation. He looks skeptical.

Scott: But I know Michael over here has been writing a lot claiming the opposite. That the modern world is terrible at coordinate problems, especially compared to the past. I'm somewhat sympathetic to that argument. In the old days, a king could just declare we were going to do something and it got done. Now we have nightmarish failures of coordination, like the Obamacare bill where the leftists had a decent and coherent vision for how healthcare should work, the rightists had a reasonable and coherent vision for how healthcare should work, and we smashed them together until we got a Frankensteinian mashup of both visions that satisfied no one. Or how back in the old days, the Catholic Church pretty much controlled…

Robin: Kings and the Church were very good at *acting*, not at *coordinating*. They could enforce their choices, but those choices were often terrible and uncorrelated with what anyone else wanted. Modern institutions *coordinate*.

Michael: But modern coordination is just through increased bureaucracy.

Robin: Call it what you want, it's still coordinating.

Michael: And the results are often terrible!

Robin: Yes, coordinating seems to divide into two subproblems. The first is getting everyone to agree on a solution. The second is making sure the solution is any good. I

don't claim we have solved the second subproblem, but we seem to be increasingly skilled at the first.

Michael: Really? Like the largest-scale world-coordinating organization we have right now seems to be the United Nations, and it's famous for not getting anything done.

Robin: The thing with the UN is that at the beginning people expected it to be the umbrella organization under which all world affairs were conducted. But there are a host of other more or less associated organizations like the WTO that are actually doing a lot more.

Scott: You make an interesting case that future coordinating power will be better, but saying "let's leave this to the future" only works if we know when the future is going to be and can prepare for it. In the case of what Eliezer calls a "foom" where an AI comes and causes a singularity almost out of nowhere – well, if we put off preparing for that for fifty years, and it happens in forty, that's going to be really bad.

Robin: I think that scenario is very unlikely. In the scenario I believe in, an increase in technology led by emulated humans, change will occur on a predictable path. They will know if we're on the path to eventual complete value deterioration.

Scott: That makes sense. So I guess that our real disagreement is only over the speed at which a singularity will happen, and whether we will know about it in time to protect our values.

Robin: Sort of. Although as I posted on my blog recently, I think "protecting values" is given too much importance as a concept. If any past civilization had succeeded in protecting *its* values, we'd be stuck with values that we would find horrible, mostly a mishmash of outdated and stupid norms about race and gender. So I say let future values drift by the same process

our own values drifted. I don't mind if future people have slightly weirder concepts of gender than I do.

Scott: I think that's kind of unfair. You're assuming the future will vary over certain dimensions where you find variation acceptable. But it might vary in much stranger and less desirable ways than that. Imagine an ancient Greek who said "I'm a cosmopolitan person…I don't care whether the people of the future worship Zeus, or Poseidon, or even Apollo." He doesn't understand that the future also gets to vary in ways that are "outside his box".

Robin: It's possible. But like I said, I think we have a very long time before we have to worry about that. I would also suggest you look at the light speed limit. That means that there's going to be inevitable "cultural variation" in the post-human world, since it will probably include a lot of semi-isolated star systems.

Scott: I still expect a lot of convergence. After all, if this is a hypercompetitive society, then they'll be kind of forced into whatever social configuration leads to maximum military effectiveness or else be outcompeted by more militarily effective cultures.

Robin: No, not necessarily. There may be an advantage for the defender, such that it takes ten times the military might to attack as to defend. That would allow very large amounts of cultural deviation from the ideal military maximum.

After this the conversation moved on to other things and I don't have as good a memory. But it was great to meet Robin in person and I highly recommend his blog to anyone with an interest in futurism or economics.

Apart from Better Sanitation and Medicine and Education and Irrigation and Public Health and Roads and Public Order, What Has Modernity Done for Us?

…

Brought peace.

As you may have noticed, instead of another GIGANTIC WALL OF TEXT I am trying to write my rebuttal to Reactionary philosophy in the form of several smaller posts that I can then link together in a sequence index. This particular post addresses Reactionary claims that modern society causes international instability, leading to increased war (or increased "total war") and the resulting mayhem.

This claim I received mostly from blog posts I can't find right now and from discussions with Michael Anissimov. It goes that when states are fully sovereign, self-interested, and run by noble classes – as they were long ago – their wars are rare, as short as possible, and mostly fought in a civilized way.

But when states are subject to a larger international order (like the UN or "international law"), interested in ideological concerns, and governed by a host of factions competing for democratic power – as they are today – wars are more common, bungled into increased length and fatality, and turn into "total war" where anything goes and civilians are considered valid targets.

Michael specifically mentioned the Congress of Vienna as an example of the old order, pointing out that a bunch of

aristocrats met up, divided Europe among them, and there was peace for decades afterwards. He compared this to the inelegance of modern "police actions" and "foreign interventions", pointing out how World Wars I and II, at the beginning of the modern era, were unmatched in their deadliness and brutality.

Luckily, these questions about war and the stability of different models of international relations can be investigated empirically. Are wars worse today, or were we worse during the old aristocratic era? By what standards?

Let's ask the media! War Is Going Out Of Style, says the *New York Times*. War And Violence On The Decline In Modern Times, trumpets NPR. Josh Goldstein says we are "winning the war on war", Steven Pinker proclaims the victory of the better angels of our nature, and John Mueller even more triumphantly posits that War Has Almost Ceased To Exist

The statistics bear them out. The BBC notes:

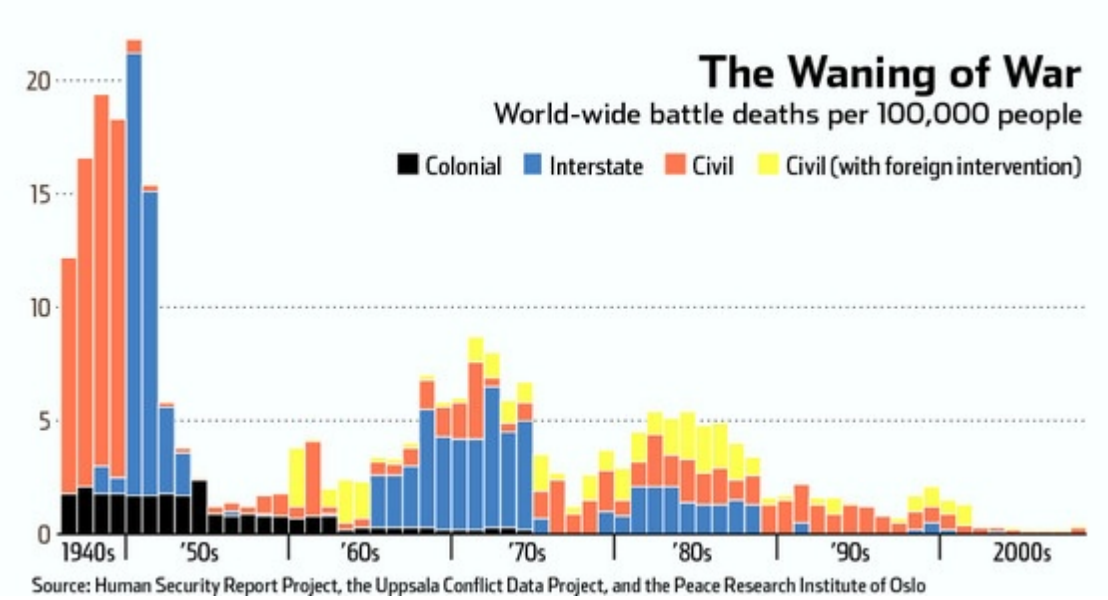
> The Human Security Report found a decline in every form of political violence except terrorism since 1992. "A lot of the data we have in this report is extraordinary," its director, former UN official Andrew Mack, said.
>
> It found the number of armed conflicts had fallen by more than 40% in the past 13 years, while the number of very deadly wars had fallen by 80%.
>
> The study says many common beliefs about contemporary conflict are "myths" – such as that 90% of those killed in current wars are civilians, or that women are disproportionately victimised. The report credits intervention by the United Nations, plus the end of

colonialism and the Cold War, as the main reasons for the decline in conflict.

The trend is older than just this decade. According to Goldstein:

> In fact, the last decade has seen fewer war deaths than any decade in the past 100 years, based on data compiled by researchers Bethany Lacina and Nils Petter Gleditsch of the Peace Research Institute Oslo. Worldwide, deaths caused directly by war-related violence in the new century have averaged about 55,000 per year, just over half of what they were in the 1990s (100,000 a year), a third of what they were during the Cold War (180,000 a year from 1950 to 1989), and a hundredth of what they were in World War II. If you factor in the growing global population, which has nearly quadrupled in the last century, the decrease is even sharper. Far from being an age of killer anarchy, the 20 years since the Cold War ended have been an era of rapid progress toward peace.

And Steven Pinker shows the following graph:



So there's more than enough data to show the world has been getting more peaceful over the past seventy years. The most plausible Reactionary response would be that this is too small a time horizon: that the horrors of progressivism should be viewed over a timescale of centuries.

First of all, this shouldn't be true. A staple of Reactionary thought is that the world *has* become notably more progressive since World War II, and a hyper-willingness to attribute anything that's declined since that period to the progressive world-view. What's good for the goose is good for the gander. Second, it is very suspicious to say that the part of the data you don't have good statistics for, and *only* that part of the data, proves your point.

But in order to address this objection more fully, I tried to get fuzzy ballpark area data on the deadliness of wars in past centuries. My methodology was to comb Wikipedia's list of wars by death toll, take all the ones with casualties of one million or greater, and organize them by era. The eras I used were 21st Century So Far, 1950-2000, 1900-1950, 1850-1900, 1800 -1850, 1700-1800, 1600-1700, 1500-1600, 1000-1500, 1-1000, and 500 BC – 1. Where casualties were given as a range, I took the center of that range, except in the Taiping Rebellion where I believe the top of Wikipedia's range is crazy high and so I took nearer the bottom; where conflicts spanned more than one era, I placed them in the one containing the majority of the conflict.

I added up total war casualties for each era, then scaled them by population using 2005 as the standard – that is, deaths were multiplied so that the new number was the same percent of the 2005 population as the original was of its own era's population. Then I divided by the length of the era to give average deaths per century during that era.

The 1900 – 1950 era indeed came on top, with 626 million projected deaths per century per 2005 population. Second place was 1600 – 1700, with 442 million. Other violent periods of note were 1850 – 1900 (326M), 1000 – 1500 (230M), and 1800-1850 (106M). There was no obvious trend related to time.

However one trend worthy of note is that the 21st Century So Far and the period 1950-2000 were *by far* the two most peaceful eras of any in the study (both at about 28M).

So the most progressive periods in history are also the most peaceful. And the Reactionaries' pet period, the 1600s when the Stuarts ruled England and the Hapsburgs were still mighty, was the deadliest age of history outside a World War. I tested what would happen if I limited the domain to Europe, and the results are much the same (with the exception of 1850 – 1900 becoming much more peaceful).

This study is actually biased against me and in favor of the Reactionaries in two ways. First, I eliminated all wars with death counts less than a million, because otherwise it would have taken *forever*. But that disproportionately eliminates pre-modern wars, since they were fought among lower-population nations – a conflict today need only kill 1/7000th of the population to make my list, but one in 0 BC would have had to kill a full 1/100 or be dropped entirely.

Second, *technology*! Two days worth of airplanes dropping bombs on Dresden in the 1940s killed more people than several long and bloody medieval crusades. More modern death counts should probably be discounted to take into account the fact that we are just way better at killing each other when we want to, even though we want to much less often. Yes, the era of World Wars saw slightly greater deaths per

population than the era of absolute monarchy in Europe. But the Allies were killing people with nuclear bombs, and the Hapsburgs were killing people with bayonets. The 17th century in particular, and the past in general, just *really really sucked*.

Some Reactionaries, intuiting this pattern, have tried to dismiss it by saying that, while progressive eras have few wars, their wars are much worse – the sort of "total war" that characterized World Wars I and II, and so rose to new levels of killing and barbarity.

But this article lists the worst conflicts of all time by percent of population killed. And you have to go to number six on the list *just to get to World War II*! World War I *isn't even on the list!* The Mongols did not kill 11% of the population of Earth in twenty-one years by not being aware you could harm civilians; the various mercenaries of the Thirty Years' War were no more innocent.

One last fact noticed in the process of going through Wikipedia's wars list: in any particular era, it is always the *least* progressive countries that are having the wars. Even the miniscule death count in the late 20th and early 21st centuries is limited almost entirely to authoritarian African countries and Islamic theocracies. In neither World War was the major conflict two democracies (by any reasonable definition) fighting one another, and at least in the latter totalitarian side deserves a disproportionate amount of blame. The bloodiest conflicts of the past few thousand years, even adjusting for population, have been in China, which is basically Reactionary Utopia with an authoritarian Emperor, a Mandate of Heaven, and strict racial homogeneity. There is a lot of debate over whether two democracies have ever gone to war (answer: it depends how true of a Scotsman you are) but this very fact

should cue you in that war and democracy are not *positively* correlated (and most likely not even neutrally correlated).

So to sum up: as the world has become more progressive over the past seventy years, conflicts and deaths from conflict have dropped precipitously. Virtually every past era was much more violent than our own, and the biases of this study probably mean they were more violent even than our numbers indicated. Every single one of the five deadliest conflicts in human history occurred before the Enlightenment, and in any given era the more progressive countries both start and participate in fewer wars than the less progressive countries.

Very likely this is due partly or mostly to economic factors – the point that no two countries with McDonalds' ever go to war is a good one. But this does not negate the fact that our current political and social system is the one that economic factors decided to set up in order to achieve their economic goals.

The Wisdom of the Ancients

Were The Victorians Cleverer Than Us?, asks a new study by Woodley et al that has gotten name-dropped in places like The Daily Mail and The Huffington Post.

Meanwhile, Betteridge's Law of Headlines continues to warn us that "Any headline which ends in a question mark can be answered by the word no."

On first glance, the paper looks solid. It investigates simple reaction time, a measure which is known to be correlated with *g*, the mysterious general intelligence which is supposedly measured (to some degree) by IQ tests. People have been experimenting with simple reaction time for over a century now, so the paper asked the relatively simple question of whether it has changed over that century. They found that it had: it had gone up, signifying a decrease in general intelligence. Their explanation was dysgenics.

People have known for a long time that high-IQ people have fewer children than low-IQ people, so it might make sense genetically to believe that each generation becomes a little dumber. This pattern has stubbornly refused to appear: instead, every generation has had significantly higher IQ than the one before, an observation called the Flynn Effect. This has been attributed to various things, including better nutrition, child-rearing, and education.

What the authors of this paper do – and it's pretty clever – is say that the Flynn Effect is an environmental increase in IQ which has hidden a simultaneous genetic decline in IQ. They try to prove it by saying environmental and genetic factors affect IQ in different ways, and that genetic factors are more

likely to affect certain features like reaction time – a pattern which is called a Jensen Effect and which is on relatively solid ground. Because they find reaction time is declining, probably people are becoming genetically stupider and the only reason we can keep having a civilization at all is because our environment is getting better – which is too bad, since our environment may have stopped doing that.

All the theory here *sort of* checks out, except for the part where they say IQ changed 15 points in a hundred years, which is just a *little* bit faster than any responsible person expects evolution to progress. People critique the idea that Ashkenazi Jews could have shifted fifteen points in *nine hundred years* on the grounds that it's too fast. So let's take a closer look at their data.

Only two of their sixteen studies come from the Victorian Era: Galton 1889 (n = 3410) and Thompson 1895 (n = 49).

Francis Galton, a brilliant Victorian scientist who was a half-cousin of Darwin, is the source of 98.5% of our Victorian reaction time data – not to mention the concept of reaction time itself, several statistical tools including correlation and standard deviation, the use of the survey in data collecting, the term "eugenics", the entire science of meteorology, hearing tests, the first study on the power of prayer (he prayed over random fields to see if the crops there grew higher; they didn't), fingerprinting, the scientific investigation of synaesthesia, and a horrible warning about how not to do facial hair.

Galton's Data A Century Later, published in 1985, tells us a little about how he gained his ground-breaking reaction time statistics. He set up a laboratory in the Science Galleries of the South Kensington Museum. There he charged visitors to the

museum three pence ($25 in modern currency after adjusting for inflation) to be measured by his instruments, a process he advertised as “for the use of those who desire to be accurately measured in many ways, either to obtain timely warning of remediable faults in development, or to learn their powers.” Over the course of nine years, he attracted about nine thousand curious individuals, three thousand of whose data managed to make it into the current meta-analysis.

His colleague in Victorian reaction-time measurement was Helen Thompson Woolley, an American psychologist who published a 1903 dissertation titled *The Mental Traits of Sex: An Experimental Investigation of the Normal Mind in Men and Women* (it was, apparently a simpler time). With an optimism bordering on the incredible, Wikipedia notes that “Before Woolley, research on sex differences was heavily influenced by conjecture and bias.”

Woolley writes of her sampling technique:

> “In making a series of tests for comparative purposes, the first prerequisite is to obtain material that is really comparable. It has been shown that the simple sensory processes vary with age and with social condition. No one would question that this statement is true for the intellectual processes also. In order to make a trustworthy investigation of the variations due to sex alone, therefore, it is essential to secure as material for experimentation, individuals of both sexes who are near the sae age, who have the same social status, and who have been subjected to like training and social surroundings. Probably the nearest approach among adults to the ideal requirement is afforded by the undergraduate students of a coeducational

> university. For most of the the obtaining of an education has been the one serious business of life.
>
> The individuals who furnished the basis for the present study were students of the University of Chicago. They were all juniors, seniors, or students in the first year of their graduate work. The subjects were obtained by requesting members of the classes in introductory psychology and ethics to serve."

She found (a finding replicated by all later studies and now considered essentially proven) that women have slower reaction times than men (interestingly, this difference does *not* correlate with IQ) – but more relevant to the current meta-analysis, she found the same generally fast reaction times as Galton.

The modern studies, keeping with the zeitgeist of the modern age, are much less colorful. I only looked into the two largest: one Scottish, the other Australian. Here's what the Scottish study says of its methodology:

> The study was originally located in the Central Clydeside Conurbation (Figure 3), a socially heterogeneous and predominantly urban region, including Glasgow City, which is known to have generally poor health. Two-stage stratified sampling was used to select subjects. For the regional sample, local government districts were stratified by unemployment and socio-economic group data from the 1981 Census and 52 postcode sectors were systematically selected from these with a probability proportionate to their population size. The same postcode sectors were chosen for all three cohorts. The sampling frame used for individuals was Strathclyde Regional Council's 1986 Voluntary Population Survey—an

> enhanced electoral register that provides details of the age and sex of all household members.3 Individuals were selected from the 52 postcode sectors within each age cohort with a systematic selection with a prescribed sampling interval from a random start.

I was getting bored by the time I made it to the Australian study, but I managed to keep my attention on it long enough to note the following sentence:

> Persons selected at random from the Electoral Roll [of Canberra] were sent a letter informing them about the survey and saying that an interviewer would contact them soon to see if they wanted to participate.

Look around you. Just look around you. Have you worked out what we're looking for yet?

That's right. The answer is **selection bias**.

Back in the Victorian Age, science was done by aristocrats and gentlemen who drew their subjects from their own social groups. There were no poor people in either study, because getting poor people to participate in an experiment would require finding some poor people, who probably smelled terrible and lived in areas where there were no good restaurants.

In the Modern Age, everyone is excruciatingly Socially Aware, and studies go out of their way to look at Disadvantaged Disempowered Disprivileged Populations so their results can serve as Cutting Social Commentary.

Galton's study population was visitors to a science museum in the posh part of London who were willing to pay him $25 to participate. Thompson's population was University of Chicago

philosophy students. The two modern studies are random selections double-checked to make sure they don't undersample the poorest sections of the population.

So, uh, congratulations, authors of this paper! You have successfully proven that the average member of the population is dumber than wealthy science dilettantes *and* students at elite colleges! Go pat yourself on the back!

In case we need more rigor: according to The National Center for Education Statistics, about 2.3% of Americans went to college in 1900. In a perfect meritocracy maybe only the smartest people would go to college, but we're not a perfect meritocracy. Would it sound about fair to say that the people in college at the time were a sample of the 20% or so of the smartest Americans?

Because the IQ of someone at the 80th percentile is 113 – that is, exactly enough to explain the 14 point IQ "drop" that Woodley et al found.

This is a little harder to do with Galton's science museum visitors. The 1985 commentary on Galton's data tells us:

> As would be expected of a group of paying testees being measured in a museum, a sizable portion of Galton's sample consisted of professionals, semiprofessionals, and students. However, as may be discerned in Tables 10 and 11, all socioeconomic strata were represented.

Tables 10 and 11 turn out to be a gold mine – I worried the records of exactly who took the tests would be lost, but as you might expect of someone who basically invented statistics single-handedly and then beat Darwin in a debate about evolution as an encore, Galton was *very good* at keeping careful data.

This site tells me that about 3% of Victorians were "professionals" of one sort or another. But about 16% of Galton's non-student visitors identified as that group. These students themselves (Galton calls them "students and scholars", I don't know what the distinction is) made up 44% of the sample – because the data was limited to those 16+, I believe these were mostly college students – aka once again the top few percent of society. Unskilled laborers, who made up 75% of Victorian society, made up less than four percent of Galton's sample!

So this discredits this meta-analysis way beyond any need for further discrediting, but since I can't help beating a dead horse…

Let's talk about race. We know that studies find white people usually have faster reaction times than black people – in fact, a lot of the voluminous and labyrinthine research on race and IQ hinges on this fact. We thankfully do *not* have to enter the minefield of trying to figure out the causes of this discrepancy (biological vs. environmental vs. social) – we can just take it as a brute fact.

What percent of Galton's 1889 science museum visitors do you think were non-white? What percent of Thompson's 1895 University of Chicago students? Approximately zero? Sad to say, non-white people were as likely to be exhibits in the science museums of the day as visitors, and according to no less a figure than W.E.B. DuBois in 1900 there were only 2600 living black Americans who had graduated college.

I looked them up some stats on the sample areas for the modern studies – 6% of Glasgow is non-white, and about 12% of Canberra. So aside from selection bias affecting intelligence

which affects reaction time, we have selection bias affecting race which affects reaction time.

May I just say how annoyed I am that I have to remind reactionary eugenicist IQ researchers, *of all people*, to pay attention to race? YOU HAD ONE JOB!

Finally, there's significant IQ differences within populations of the same race and country simply due to migration effects. An analysis of IQs across Great Britain finds that the highest scores are in London (102) and the lowest in Scotland (97). Almost all this meta-analysis' Victorian data came from London (Galton's museum in Kensington) and the largest source of modern data (making up about half of the whole, and being unusually high in reaction time) came from Scotland (and Glasgow isn't even the nice part of Scotland). The 5 point London – Scotland difference explains over a third of the "difference between Victorians and moderns" found in this study.

So in conclusion, this study ignores race, ignores regional variations, but most importantly IGNORES THAT ALL ITS VICTORIAN STUDIES WERE SAMPLING FROM THE SMARTEST 20% OR SO OF THE POPULATION AND THEY GOT EXACTLY THE NUMBERS YOU WOULD EXPECT IF YOU DID THAT.

There is some *really excellent* IQ research out there that everyone should be reading, but this is not it. Please please *please* don't cite this study as evidence for dysgenics or the decline of civilization.

Can Atheists Appreciate Chesterton?

Empirically, yes.

Friday was the anniversary of Chesterton's death, the religious blogosphere is eulogizing him, and I thought I'd join in. I enjoyed and recommend Chesterton's novels, especially The Man Who Was Thursday and Napoleon of Notting Hill, his works of nonfiction like Heretics, and even his poems (all of these are links to freely available fulltext versions online).

Classical philosophy holds that evil is merely the absence of good, but for me, at least, the opposite reduction is more tempting (albeit just as wrong). Evil is extremely obvious – you can look at people involved in animal cruelty, or bullying, or whatever, and you can almost *see* the actively malicious force animating them onward. On the other hand, good is most easily perceived as unusual skill at avoiding evil. Vegetarians are unusually good because they take extra effort to avoid hurting animals, people who donate to charity are unusually good because they take extra effort to avoid greed.

I credit three authors with giving me a visceral understanding of active, presence-rather-than-absence Good: G. K. Chesterton, C. S. Lewis, and Jacqueline Carey. Two of those are very religious and write quite consciously from a Christian perspective. The third writes about kinky sex. Go figure.

But actually when I think about it more closely, the moral beauty in Carey's writing comes mostly from her constructed religion, which is *suspiciously* similar to Christianity. So it seems that there's a fact to be explained here.

Can an atheist appreciate Chesterton? A better question might be whether an atheist can happily appreciate Chesterton as

offering a beauty that she, too, can partake in, or whether the appreciation must be along the lines of "Yup, these are the nice things we can't have."

Keep The Horse Before The Cart

So I think an important point to make before going any further is that, through 90% of Christian history G. K. Chesterton and C. S. Lewis probably would have been burnt at the stake.

Not just for denominational reasons, although that would have been enough. Promoting joy as a sign of sanctity and as a proper state for man – that's a burning for the Epicurean heresy right there. Believing righteous non-Christians could get into Heaven – that's a burning. A suggestion that that humor and lightness were chief attributes of God and the angels – more burning. Doubting the literal truth of some of the Old Testament? Uncertainty whether the New Testament was divinely inspired in a more-than-metaphorical all-great-art-is-divinely-inspired way? Claims that praying sincerely to false gods was praiseworthy and basically just another way of praying to God? Burning, burning, *burning*.

The moral qualities that shine in Lewis and Chesterton – joy, humor, a love of the natural world, humanity, compassion, tolerance, willingness to engage with reason – are all qualities they inherited from modernity which would be repugnant to many of their Christian predecessors. They are all totally within the milieu of early 20th century England and totally foreign to medieval Italy or ancient Judea.

St. Augustine could not have written *The Great Divorce*, because while Lewis was talking about how the blessed in Heaven suffer great hardship to meet the damned in order to radiate love and wisdom at them and help bring them to Heaven, Augustine was writing about how the greatest

pleasure of the blessed was getting to watch the tortures of the damned, metaphorically munching popcorn as they delighted in sinners getting what they deserved. Tertullian didn't even wait until after he died to start getting delighted, famously saying that:

> "At that greatest of all spectacles, that last and eternal judgment how shall I admire, how laugh, how rejoice, how exult, when I behold so many proud monarchs groaning in the lowest abyss of darkness; so many magistrates liquefying in fiercer flames than they ever kindled against the Christians; so many sages and philosophers blushing in red-hot fires with their deluded pupils; so many tragedians more tuneful in the expression of their own sufferings; so many dancers tripping more nimbly from anguish then ever before from applause."

What Lewis, Augustine, and Tertullian had in common was Christianity; what set Lewis apart was modernity. What made C. S. Lewis saintly, as opposed to the horrifying sadists who actually got the "St." in front of their names, was the perspective of a culture that had just spent a few centuries thinking about morals from a humanistic perspective.

When Pope Francis said that we need to build a "culture of life" that can protect innocent children from harm, he wasn't taking a revelation from the Biblical angels but from the Better Angels Of Our Nature. The *Biblical* angels are the ones who would be tasked with enforcing God's promise of blessing on anyone who takes Babylonian infants and smashes them against rocks (Psalm 137:9, look it up).

During the tradition from the Dark Ages to modernity, people got technologies like the printing press and the frigate and started learning more about other cultures, seeing that they

were decent people and that no one religion had a monopoly on morality. The decline in infectious diseases banished death from an everyday presence to a lurking evil and made casual slaughter seem less appealing; the gradual decline in war resensitized people to violence. And all this time there were philosophers inventing things like deontology and consequentialism and freedom and equality and humanism and saying that yes, people did have inherent moral worth. And religion eventually decided that if it couldn't beat them it might as well join them, at least to a degree, and it was this concession that allowed the moral decrepitude of people like Tertullian and Torquemada to evolve into the moral genius of people like Chesterton and Lewis.

So my thesis is that Lewis and Chesterton didn't become brilliant moralists by revealing the truths of Christianity to a degraded modern world. They became great moralists by taking the better parts of the modern world, dressing them up in Christian clothing, and handing them back to the modern world, all while denouncing the worse parts of the modern world as "the modern world".

And so rah humanism and all that. But the original question remains: what is it about the Christian clothing that is such a necessary ingredient?

A Cupboard Full Of Secret Ingredients

First of all, the power of myth.

I don't think it's a coincidence that all three of the people I named as influences on my sense of moral beauty were writers of speculative fiction. Fiction has greater opportunity to be beautiful and to show complicated internal dynamics of humanity than abstruse philosophy or dry preaching does, and speculative fiction has a better opportunity to present

superstimuli, including moral superstimuli. I think that people who write speculative fiction ordinarily tend to be kind of dismissed, but that because Lewis and Chesterton were working from within a tradition that had its own myths, they managed to get through the filter of "Oh, it's just fantasy, ignore it". Narnia was dignified by being a metaphor for the Bible, which earned its dignity through hoary age and civilizational influence.

Second of all, legitimacy.

I sometimes write about morality. It tends to be in a light-hearted "here's what I think" style, first of all because I'm genuinely uncertain about a lot of stuff, second of all because I don't want to sound preachy. Religion is really good at helping people be certain of things, and religious people get a free pass to sound preachy because preaching is what religions are *supposed* to do.

I don't think there's a niche for non-religious versions of Chesterton and Lewis. There are people like that New York Times ethics columnist who talk about ethics, but I think if they were to start getting *poetic* about it, people would start challenging their right, be like "Who told *you* what is or isn't necessary for the integrity of the human spirit?" This is a tough question. But Lewis and Chesterton have a great answer: "God did". They can, as the Bible puts it, "speak like one who has confidence".

Third of all, a different perspective.

You can seem deep just by saying something different than everyone else does. I don't think Lewis and Chesterton were too far from the modern moral mainstream, but I think they use a completely different aesthetic. Where most people talk about the bravery of defying the mainstream, a Christian

writer can talk about the bravery of *not* defying the mainstream when everyone thinks you should. Where most people talk about the importance of high self-esteem, a Christian writer can talk about taking care to avoid pride. Both sides have valid and important insights, but if a culture is doing everything it can to saturate you with one of them, the other will be a powerful breath of fresh air.

Chesterton – I haven't yet noticed this in Lewis – has this sort of gambit where he agrees with some modern virtue, and then says the correct way to attain the modern virtue is through doing the opposite of the modern virtue. Or maybe the opposite, where he agrees with what we should be doing, but then says the end goal is exactly the opposite of what everyone would think:

> The outer ring of Christianity is a rigid guard of ethical abnegations and professional priests; but inside that inhuman guard you will find the old human life dancing like children, and drinking wine like men; for Christianity is the only frame for pagan freedom.

People make fun of this, and rightly so (Steven Kaas attributes to Chesterton's dog the quote "Arf arf arf! Not because arf arf! But exactly because arf NOT arf!") but I think it is fundamental to his project. He gets to maintain his belief in modern virtues while getting there through an unexpected path that seems deep and profound and unexpected.

Fourth of all, a focus on the individual.

Despite everything everyone says about modern society being too individualistic, there seems to be a sense in which the opposite is true. The problems we are comfortable talking about are ones like racism, sexism, income inequality,

terrorism, crime. Social problems. Problems in the community. The idea of talking about what goes on in the individual soul, of having strong opinions about it, isn't a very modern sensibility at all. The only exception are psychologists and therapists, who really want to be scientific and so scrupulously avoid sounding poetic.

I could come up with some just-so stories about why this is – we like to think scientifically, but intrapersonal dilemmas don't lend themselves to this kind of analysis? Focus on individuals doesn't generalize well, which is a problem in the age of mass media? Christians were abnormally obsessed with the individual soul because of virtue ethics + the idea of damnation and salvation? I'm not sure. Anyway, religion has a head start on individualist vocabulary and thought processes which non-religion doesn't really have good alternatives for (PSYCHODYNAMICS DOES NOT COUNT AS A GOOD ALTERNATIVE).

All of these are kind of banal and not the sort of thing that could prevent an atheist from fully appreciating Chesterton. But then there's the big one.

What Lewis, Chesterton, and Carey have in common is this belief in Good as an active, vibrant, force, in Good being not just powerful, but so powerful that it's kind of terrifying. As something not just real, but *the most real* thing.

Atheists can have Good be terrifying – utilitarianism has broken much stronger minds than my own – but it's really hard to have it be *real*. I'm not saying atheists can't believe in Good, just that atheist good is a sort of – I hate this term but I'll use it anyway – social construct. It's real in the same sense the US Government is real. The US Government is certainly powerful – just ask any Iraqi. But it's not *one thing*, with an

essence and a personality and angel wings of red-white-and-blue fire. It's just an abstraction over a lot of ordinary people doing their thing.

And this would seem to be the death blow for atheists having something as strong and convincing as a Lewisian or Chestertonian world-view. Except that I kind of picked up a similar vibe from *Harry Potter and the Methods of Rationality*. I didn't think of it when I was naming the three authors who first made me think of Good as a thing, but it is another work that portrays Good as this burning, all-powerful force, and although it has some magic in it, it doesn't go all the way to reinventing Christianity like Carey does.

I'm not sure whether this is sleight-of-pen, whether it only works because of the magic there because even if the magic and morality aren't explicitly linked it still triggers sort of morality-is-magic circuits. Or whether it only works if you're literally responsible for saving the world. But it seems encouraging.

I think the truth of Lewis and Chesterton is not only appreciatable by atheists but derives from humanist ideas. The *beauty* of Lewis and Chesterton I'm not sure about, but I maintain some hope that it can be saved as well, even if I'm not sure how to do it.

Holocaust Good for You, Research Finds, But Frequent Taunting Causes Cancer in Rats

A study published this month in PLoS One finds that victims of weight discrimination ("fat-shaming", in case you only speak Tumblrese) are more likely to subsequently gain weight.

It's hard for me to like a study that so obviously got exactly the result its organizers wanted it to get. And obvious confounders are obvious – level of discrimination faced was based on self-report, and the sorts of people who hang around the sorts of people who fat-shame may differ systematically (in class? education?) than who avoid that kind of abuse – but the study's endpoint of *change* in weight over time rather than just weight itself goes some of the way toward addressing those concerns. And I've got to give them credit for studying an important issue and getting a highly significant result. So let's let them have their soapbox:

> There are both behavioral and physiological mechanisms that may contribute to the relation between discrimination and obesity. Weight discrimination is associated with behaviors that increase risk of weight gain, including excessive food intake and physical inactivity. There is robust evidence that internalizing weight-based stereotypes, teasing, and stigmatizing experiences are associated with more frequent binge eating. Overeating is a common emotion-regulation strategy, and those who feel the stress of stigmatization report that they cope with it by eating more. Individuals who endure stigmatizing experiences also perceive themselves as less competent to

> engage in physical activities and are thus less willing to exercise and tend to avoid it. Finally, heightened attention to body weight is associated with increased negative emotions and decreased cognitive control. Increased motivation to regulate negative emotions coupled with decreased ability to regulate behavior may further contribute to unhealthy eating and behavioral patterns among those who are discriminated against.

New study! This one published – oh, look, isn't that interesting – this month in PLoS One, finds that survivors of the Holocaust have greater life expectancy than control Jews who did not experience the Holocaust.

Here the authors definitely got a result they were *not* looking for and did *not* want. And here, too, we have all sorts of confounders: they tried hard to construct a matched control group of Jews who emigrated from Poland to Israel just before the Holocaust, but we have no idea what sort of differences there might have be in those populations (just to make up one story, maybe poor people who had less to lose were more likely to emigrate). And here too, there is no shortage of soapboxes. From here:

> One possible explanation for these findings might be the "Posttraumatic Growth" phenomenon, according to which the traumatic, life-threatening experiences Holocaust survivors had to face, which engendered high levels of psychological distress, could have also served as potential stimuli for developing personal and inter-personal skills, gaining new insights and a deeper meaning to life. All of these could have eventually contributed to the survivors' longevity. "The results of this research give us hope and teach us quite a bit about

> the resilience of the human spirit when faced with brutal and traumatic events", concluded Prof. Sagi-Schwartz.

So, let me sum up what we've learned here today.

Having someone call you fat is a profoundly disturbing form of stigmatization that breaks your normal cognitive coping mechanisms and subjects you to levels of stress that the human body and psyche were never designed to withstand.

But being rounded up like cattle, having your entire family killed in front of you, and then being starved nearly to death in a concentration camp for several years is useful opportunity to grow as a person, and will leave you stronger and better-adjusted.

I shouldn't be too sarcastic. Stranger things have ended up being true. Maybe constant low-grade minor stress has a deleterious effect but a single extremely stressful event can be salutary. Maybe stress is good for you only after you've achieved a safe distance from the stress and can reflect on it from a position where you're absolutely sure it will never happen again. Maybe stress makes you obese in the short term, but also makes you live longer in the long-term. Maybe the cultural differences between elderly Polish Jews and middle-aged Americans mediate the effect stress has on their bodies.

Or maybe these effects are mediated by unexpected processes. Maybe the Holocaust survivors live longer not because of personal growth, but because they got a sort of involuntary caloric restriction that permanently altered their metabolism. Maybe (as the researchers point out in their paper) only people who were exceptionally healthy survived the Holocaust, and these people continued being exceptionally healthy into their old age. Maybe obese people who aren't shamed stick to a

careful diet to avoid shaming, but once the shaming starts they figure it can't get any worse and go wild.

Or maybe one or both of these studies is totally and fundamentally flawed and we're wasting our time here. I give 50% probability that the fat result is legitimate, and 90% probability the Holocaust result is due to something other than personal growth, probably survivor effect or caloric restriction – but I bet others will disagree.

Yet I think what struck me most about this combination was how "stress makes you miserable and unhealthy" sounds reasonable, but "stress is a salutary process that allows you to grow" also sounds reasonable. No matter what happens to stressed people, psychology can go "Oh yeah, according to our theories, stress causes that" and I will nod my head and agree.

Or maybe another way to put it is that I'm impressed with the ease at which we switch narratives. *All the time* I hear "Well, a little bit of adversity will be good for him/her". Or else "What you're doing is going to destroy his/her self-esteem and scar him/her for life." Most people selectively use either or, depending on whether they want to excuse something or condemn something at that particular moment, and they have the science availablecached thoughts, but we have a store of contradictory cached thoughts sufficient to support any proposition *or* its opposite.

This is why the Ethics Committee needs to hurry up and approve my replication experiment to commit genocide against a randomly selected sample of the population.

Public Awareness Campaigns

A little while back I discussed how, contrary to the conventional wisdom that they've been "proven to work", anti-rape campaigns aimed at men have zero evidence of effectiveness. I added that this was no surprise, since similar public awareness campaigns have a long history of failure.

That was overly simplistic, as commenters quickly reminded me. Some public awareness campaigns (or things like public awareness campaigns) have a history of spectacular failure. Others have a history of spectacular success. To give a couple of examples:

Failures

– DARE (Drug Abuse Resistance Education) is a program in US schools where teachers and police officers spend a couple of hours a week telling children how bad drugs are and giving them techniques to avoid peer pressure to use them. It famously fails to work and in some cases even makes children more likely to do drugs.

– Scared Straight is a popular program in which convicted inmates talk to deliquent kids and explain the costs of a life of crime and how unpleasant prison can be. Several studies clearly show that the intervention makes these children actively more likely to become criminals, and is so harmful that "each dollar spent on Scared Straight programs incurs costs of $203.51". Needless to say there continue to be dozens of these programs all around the country.

– Sex education in schools is famously ineffective. There is a very large body of research showing abstinence-only sex education programs do not make teens less likely to have sex.

Research on comprehensive (ie contraception-including) sex education programs is more mixed, and although everyone trumpets the positive results in order to further discredit the abstinence-only programs by comparison, the actual research is more nuanced and less optimistic. The limited effects it does get may work by spreading genuinely novel information ("condoms exist! STDs exist!") rather than "awareness raising" per se.

– "Diversity training" and "sensitivity workshops" and everything in that category of thing have no positive effects are are often associated with negative effects (for example, after being introduced in companies those companies' percent of top executives who are minorities goes down)

– A couple of recent studies (1, 2) are converging on the hypothesis that stigmatizing overweight people makes them *more* likely to gain weight. This is true even when the stigma is delivered in the form of a (presumably more polite) public awareness campaign instead of just an acquaintance calling you a lardass. Although I agree a line can be drawn between "public awareness campaign" and "stigma", in practice it can sometimes be kind of fuzzy – for example, although most people wouldn't use the word, it sure *seems* like the point of "Don't Be That Guy" anti-rape campaign is to stigmatize rape.

Successes

– Several people have brought up MADD's campaign against drunk driving, which corresponded to a 65% decrease in drunk driving over the past 30 years. However, I can't find good evidence on whether MADD started a traditional public awareness campaign or just lobbied for changes in various laws and got lots of publicity in the process. I would also note that there have been spectacular and somewhat mysterious

decreases in nearly all crimes since 1982 – alternately attributed to rising abortion rates, falling lead rates, and stricter sentencing – and it's not obvious how much drunk driving is just piggybacking on that success.

– Seat belt use has gone from very low to near-complete, and there is decent evidence that awareness campaigns like Click It Or Ticket contributed to this (fun fact: opponents of mandatory seatbelt laws launched a counter-campaign called "Stick It To Click It Or Ticket").

– Advertising is kind of like a "public awareness campaign" about a particular product. It obviously works or else companies wouldn't spend so much money on it.

– Anti-smoking campaigns do seem to lead some people to stop smoking – or at least increase calls to stop-smoking hotlines. These are most effective when associated with scary and graphic images – for example, one shows a picture of a man with a hole in his throat after a throat cancer operation. There are a *lot* of successful public health campaigns along these lines.

Analysis

It's pretty hard to draw a consistent "this works, that doesn't" conclusion from these facts.

Just to give an example, one of the most effective campaigns – anti-tobacco – uses the same strategy as one of the least effective campaigns – Scared Straight. Both try to present very graphic images of the horrible things in store if people do not change their ways. One works great, the other is counterproductive.

To give another, both anti-obesity and anti-drunk-driving campaigns try to employ stigma, but one of them has been

very successful and the other has if anything the opposite of the intended outcome.

Some people I talked to about this at the New York Solstice Celebration suggested that product advertising works because businesses have financial incentives to get it right, but other public awareness campaigns don't because the government mostly wants to signal virtuous effort and has no incentive to design genuinely effective advertising change minds. But some government and nonprofit public awareness campaigns are successful, and I'm betting they're all hiring the same ad agencies anyway.

Another theory was that awareness campaigns work when there's a real need for awareness – either the target demographic is literally unfamiliar with the concept in question (for example, people may not have previously been aware the police were cracking down on non-seatbelt-users) or need reminders to keep an option fresh in their minds (Coke advertisements making everyone think about Coke when they're deciding what to buy). If it's just stating *ad nauseum* that some stigmatized action like premarital sex is still stigmatized, it's not going to do much. But this is disproven by the success of MADD and the stop-smoking campaign, and by the failures of DARE (which very often does teach kids things about drugs they didn't know before).

The biggest effect I can see is that anything which caters to a captive audience is more likely to be counterproductive. DARE and sex-ed are inflicted on schoolchildren who would rather be doing something else, and a lot of the time it ends up as "this uncool authority figure I don't like lectures about how me and all my friends are bad people". Scared Straight programs are usually court-mandated, often as a punishment for past delinquent behavior. Employees are forced to attend

diversity training, and once again it may be billed as a "punishment" for saying something politically incorrect. Is it so far-fetched that people forced to suffer through these campaigns will end up resentful, and that resentment will translate into negative feelings about the campaign message?

I'd like to extend the theory to the obesity case and say that, once again, stigmatizing the obese in anti-obesity campaigns causes obese people to associate the negative feelings they get from these campaigns with "eat less and exercise" message. But this proves too much: why wouldn't the scary disfigured people in the stop-smoking ads make smokers associate their negative feelings with quitting? Perhaps the negative feelings have to be of a certain type for this to work? Anger and resentment, rather than fear and disgust? Questions, questions.

I still don't feel like I have a good ability to predict the success or failure of any future public awareness campaign. If I wanted to promote the "Don't Be That Guy" anti-rape campaign, I would point out that it consists mainly of flyers on lampposts etc, so there's no captive audience nor any reason to consider it a "punishment". If I wanted to inveigh against it, I'd argue that empirically it offends a whole lot of men who think they're being binned as potential rapists and so definitely causes the anger and resentment which are the hallmark of a counterproductive campaign. I really don't know.

I guess part of the reason I remain skeptical of public awareness campaigns is a lingering terror at what it would say about society's collective sanity if they worked. Think about it. Imagine that TV ads warning people not to do drugs *really* decreased drug use. Then think about how for the past 30 years, we haven't been consistently running those ads, but we *have* been consistently putting anyone who uses drugs behind bars for their entire lives. Imagine if anti-rape ads worked, and

only two Canadian cities have ever run them. And *no* city afaik has ever run ads against child abuse!

I would welcome more examples of public awareness campaigns that clearly succeeded or failed. Post them in the comments. Please exclude ones that measure "success" by surveying people about whether they saw the campaign or became aware of the campaign message – I'm interested in ones that actually change behavior.

Social Psychology is a Flamethrower

Mark Twain:

> There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact.

If this is true of all science, it is doubly true of social psychology.

At its best, social psychology is an unmatched window into human motivations, a "look under the hood" of the way people talk and act. The best research in social psychology is as well-supported as anything in physics or biology, and much more intuitively comprehensible. This is why it's one of my favorite scientific fields.

But at its worst, social psychology is a flamethrower. People grab hold of it to try to fry their political opponents, then end up lighting their own hair on fire or burning down half a city. Because social psych is *really hard* to do right.

Social psychology experiments in the laboratory tend to throw up spectacular mind-boggling effects. Many of these fail to replicate and are later discredited. The ones that do replicate are not always generalizable – sometimes an even *slightly* different situation will remove the effect or create exactly the opposite effect. The effects that remain robust in the laboratory may be too short-lasting or too specific to have any importance in real life. And the ones that do matter in real life may respond unpredictably or even paradoxically to attempts to control them.

This is relevant because a lot of our political discourse revolves around ideas lifted from social psychology. Every time someone advocates banning violent videogames so that they don't normalize violence, they're using social psych. Anyone who says the media needs more positive role models of minority groups and fewer stereotypes, they're taking terms out of the social psych lexicon. Whenever you complain that magazines objectify women, you're implicitly buying into several social psych theories.

Most people are not consequentialists, but most people feel implicitly uncomfortable making moral arguments on non-consequentialist grounds. "Stop what you're doing, it disgusts and offends me" is less noble than "stop what you're doing, it will hurt people who can't stand up for themselves". This tempts people who are disgusted and offended by things to come up with just-so stories from social psychology for why the disgusting and offensive thing will also hurt people.

I tried writing a post arguing against several of these just-so stories, but it ended up being *unbearably* long and boring (if you're ever stuck with insomnia, ask me to give you a trenchant analysis of every study that's ever been written about stereotype threat). So I'm going to try something different. I'm going to write up some just-so stories using social psychology for the opposite side. I'm going to try to use well-established social psych results to prove that we should have *more* violence in the media, and be *more* tolerant of offending women and minorities.

I think some of the arguments below will be completely correct, others correct only in certain senses and situations, and still others intriguing but wrong. I think that modern pop social psychology probably contains the same three categories

in about the same breakdown, so I don't feel too bad about this.

Violence In The Media Prevents Violent Crime

Dahl and DellaVigna (2008), well aware of laboratory experiments that found violent media temporarily made subjects more violent, decided to investigate whether the opening weekends of blockbuster violent movies affected crime rates. Sure enough, they found they did…

…in the opposite of the expected direction. They found violent movies decreased crime 5% or more on their opening weekends, and that each violent movie that comes out probably prevents about 1000 assaults. Further, there's no displacement effect – the missing crimes don't pop back the following week, they simply never occur.

They hypothesize that every hour violent criminals are at the kind of movies that appeal to violent criminals is one hour more they're not getting drunk or taking drugs or committing violent crimes. Although they don't mention it directly, other analyses have suggested that the movies have a sort of cathartic effect, satisfying their urge for violence without them having to commit it themselves.

An investigation into violent video games found essentially the same pattern: violent video games decrease crime while nonviolent video games have no effect.

There are also studies that show that playing lots of violent video games is correlated with violent criminality, but a much more plausible explanation of the data is that a naturally violent personality makes people more likely to enjoy violence both in games and in real life.

Decreasing violence in the media might therefore be predicted to increase violent crime, both by putting more criminals out on the streets and by sabotaging their attempts to indulge their violent urges in an acceptable manner.

Media That Objectifies Women Prevents Rape

Just as violent movies prevent violent crime, pornography may prevent rape. It's easy to prove that in the US every 10% increase in Internet access causes a 7.3% decline in rape, and it's not due to any of the expected confounders. Another study points out a similar correlation in Japan. I find the particular correlation they mention very sketchy, but Japan does have a very low rate of reported sex crimes (a commenter brings up the possibility that Japanese culture merely discourages reporting). Other more rigorous studies on the Czech Republic show the same, and studies on child porn show pedophilia is less common where it's more accessible. And these studies links to more interesting results, mentioning how sex criminals are less likely to consume pornography than the general population and start watching pornography at a later age.

This is explicable not only by the substitution effect mentioned above, but by the general tendency of orgasm to relieve frustration. If, as has been hypothesized, rape is an expression of anger and powerlessness at the world in general or women in particular, orgasming to violent porn is going to both satisfy that aggressive impulsive and replace it with general post-coital relaxation.

Saying Tests Are Biased Against Minorities Makes Minorities Perform Worse On Tests

It is relatively clear that achievement gaps on standardized tests – black-white, male-female, and the others – are not due to bias in the tests themselves. Although some sociologists

raise the specter of "tests that claim to be fair by asking both rich and poor people the same questions about golf and yachting", in real life achievement gaps remain mostly consistent across verbal tests, pure mathematical tests, symbol manipulation tests, and extremely basic and un-bias-able tests like ability to remember numbers backwards.

This has not stopped the constant repetition that various specific tests – SAT, GRE, IQ – are biased against minorities.

We know exactly what happens when minorities are told tests are biased against them: they do worse on those tests. This is the essence of the idea of "stereotype threat" – for example, one can improve women's performance on a math test simply by telling them that the test is not biased against women. So maybe we should stop doing exactly the thing that we just proved hurts women and minorities' educational performance.

Fighting Stereotypes Makes People More Prejudiced

The largest-ever study on diversity training, following 830 large companies over 31 years, found:

> A comprehensive review of 31 years of data from 830 mid-size to large U.S. workplaces found that the kind of diversity training exercises offered at most firms were followed by a 7.5 percent drop in the number of women in management. The number of black, female managers fell by 10 percent, and the number of black men in top positions fell by 12 percent. Similar effects were seen for Latinos and Asians.

Similarly, all studies on sensitivity training find that trainees express more *awareness* of sexual harassment than non-employees, but a study that went further and examined results found that trainees are "less likely to perceive coercive sexual

harassment, less willing to report sexual harassment, and more likely to blame the victim".

This is not particularly unexpected: we know for example that nearly every study on DARE programs has found that they increase drug use, sometimes as much as 30%.

Why should this be? Three reasons come to mind. The first is a boomerang effect from the programs themselves. Diversity training, sensitivity training, and DARE are all things busy people are required to attend where they (essentially) are forced listen to people behave condescendingly to them. This makes them dislike the training, their instructors, and, by association, the opinions they are trying to get trained into them.

A second reason is more fundamental. The backfire effect is when people challenged with information that disproves a cherished political belief of theirs react by becoming even more certain of the belief. The link will fill you in on potential explanations.

And the third reason is what the Harvard Business Review Blog, in its discussion of the diversity training study above, described as "when people divide into categories to illustrate the idea of diversity, it reinforces the idea of the categories."

I'll admit I had a sheltered upbringing and may be atypical, but I would estimate about 90% of the racist stereotypes I have ever heard were part of efforts to fight racism. No one just comes up to you and says "Hey, you know black people? Pretty unintelligent, huh?" (at least not to *me*). But social justice people will repeat the stereotype about black people not being intelligent again, and again, and again, to anyone who is anywhere near them, in the guise of fighting it.

I can't find the link for this, but negatively phrased information can sometimes reinforce the positive version of that information. For example, if you tell people "President Obama is not a Muslim", then a year later, all someone will remember is "blah Obama blah blah blah Muslim", and eventually "Ohmigod, President Obama is a Muslim!", even if they didn't believe that before they heard that fact "corrected".

Imagine I told you "People from Comoros are *not* all homosexual! This is a damn lie, and anyone who says people from Comoros are homosexual is an insensitive jerk. Please join me in fighting the popular perception that everyone from Comoros is a flaming gay.

Go ahead, try to think of Comoros in *any context* other than an archipelago full of gay people now. I'll wait. Take a whole lifetime, if you want. It won't help. Ten years after this blog is deleted and this post is inaccessible except through archive.org, there will still be a couple dozen people who are convinced that everyone from Comoros is gay, because they "heard it somewhere". At the very least, the idea of Comoros = homosexuality is now firmly implanted in your mind, and it will be impossible to meet a Comorosian without secretly evaluating her sexual orientation and then trying to stop yourself from doing it.

Now imagine instead of hearing this once, you heard it every day of your life.

Calling People Racist Makes Them More Racist

Foster & Misra (2013) is a jewel of a paper I stumbled across totally by chance.

They got a bunch of undergraduate students in romantic relationships and gave them a test that asked them some questions about infidelity – things like "is it unfaithful to

fantasize about another girl/boy when you're in a relationship?". They pretended to grade the test, but in fact they ignored the test and gave fake feedback.

The control group was told that they had some of the highest faithfulness scores of anyone in the experiment, they must be really faithful, good job. The experimental group was told they had some of the lowest faithfulness scores of anyone in the experiment and that the test had pegged them as having an unfaithful personality type. Once again, all this feedback was fake and both groups got around the same average score.

Then they measured what they called "trivialization" in both groups – that is, they asked them questions about how important faithfulness was to them. Consistent with their theory, the people who were told they were faithful said faithfulness was extremely important, but the people who were told they were unfaithful "trivialized" the behavior – who cares about fidelity anyway, infidelity is *maybe* a minor mistake but it doesn't really hurt anyone, people should really stop whining about infidelity all the time. To give you a feeling for the size of this effect, on a scale of one to seven, the faithful group rated the importance of being faithful at 5.4/7, and the unfaithful group rate the importance of being faithful at 2.9/7. In other words, by accusing them of being unfaithful, the experimenters had successfully gotten the participants to "trivialize" faithfulness.

The researchers theorized that this was the process called "cognitive dissonance". Most people like themselves and want to continue to like themselves. If they are told that they, or their group, has a particular flaw, then instead of ceasing to like themselves it may be easier to just decide that flaw is not a big deal and they can have it while continuing to be the awesome people they secretly know they are.

Now not only do the experimental subjects here stop caring about being faithful, but everyone pushing a pro-fidelity line is a threat to their new identity. And *the subjects weren't even really unfaithful to begin with!*

Modern political discourse tends to do a lot of things like say "All white people are racist" or all men are naturally prone to violence and potential rapists. Or it may take little things normal people do and tell them they are racist or creepy or rape-y or something because of it.

What this does is drive people into identifying with these negative labels. And instead of making them want to change their behavior to stop identifying with these labels, it may just make them think "Well, if *I* do it, then I guess it can't be so bad."

Talking About Rape Culture Causes Rape

There is a strong debate still going on about whether the death penalty decreases crime. But this hides a more settled question, which is whether punishment decreases crime at all. The relatively accepted answer is yes, it does.

Criminologists have tried to separate out the important of punishment into two aspects: severity and certainty. They have consistently found that the certainty of the punishment is more important than the severity – the most important factor in whether someone commits a crime is the likelihood she will be punished.

No criminal can see into the future to discover whether or not they will be punished; the only way certainty of punishment can influence crime is through public perception of certainty of punishment. That suggests that if you discover that an abominable crime has (contrary to popular perception) a very

low chance of punishment, it would be an excellent time to practice the virtue of silence.

Or consider the claim that rape jokes cause rape. As I understand it, the claim goes that someone tells a rape joke, then everyone else laughs, no one protests or anything, and then potential rapists in the audience conclude that they are in a culture that considers rape acceptable.

You know what else could potentially cause people to think our culture considers rape acceptable? Writing and publicizing countless books and articles arguing elegantly and vehemently for the point that our culture considers rape acceptable. Seriously. If I were a demon from Hell, charged by my infernal masters with increasing rape as much as possible, I literally could not think of a better strategy than talking about rape culture all the time.

Getting angry at the rape jokes while enthusiastically taking part in the demonic campaign thing seems like (to mix metaphors) missing the mountain for the molehill.

Summary

In this post, I've give six social psychological just-so stories: media violence prevents crime, objectification of women prevents rape, accusations of test bias hurts minorities, fighting stereotypes makes people more prejudiced, calling people racist makes them more racist, and talking about rape culture increases rape.

These can be easily compared to six much more common social psychological just-so stories: media violence causes crime, objectification of women causes rape, accusations of minorities doing worse on tests for intrinsic reasons like their culture hurt minorities, fighting stereotypes makes people less

prejudiced, calling people racist shames them out of their racism, and making rape jokes increases rape.

I don't consider any of my six completely proven, just intriguing and intuitively plausible. And of course, there's an element of concern-trolling in all of them.

But I don't consider any of the second six completely proven either; again, they are merely intriguing and intuitively plausible. And they have their own element of being suspiciously congruent to the political beliefs of the people who push them, as if they're trying to come up with consequentialist justifications for ideas they hold for other reasons.

Some will point to various studies conducted on one or another of them, but with very few exceptions all those studies have been poorly replicated investigations into the very-short-term (less than ten minutes) effect of laboratory interventions on proxy variables. These can be diametrically opposite their real social effects – for example, the laboratory experiments that experimental exposure to violence causes people to play contrived games in a more aggressive manner couldn't catch that in the real world, violent movies decrease crime. And poorly replicated short-term laboratory interventions on proxy variables can prove nearly anything – see for example the recent controversy around whether the word "Florida" makes people walk more slowly.

The six stories above suggest some pretty radical and unpalatable action approaching social engineering. For example, the idea that research into test bias should be suppressed, even if it is scientifically rigorous, just because hearing about it might hurt women – seems pretty unfair (same with the idea that no one should be allowed to talk about rape

culture) And it seems unreasonable to ask people to constantly watch their language around white people to avoid anything that sounds like accusing them of racism because that could have unpredictable negative effects on them down the line.

But the six traditional stories also suggest pretty radical and unpalatable action approaching social engineering. For example, the idea that research into gender differences should be suppressed, even if it is scientifically rigorous, because hearing about it might hurt women. Also unpopular is the idea of constantly having to watch your language around minorities to avoid anything that sounds like you're saying something racist because that could have unpredictable negative effects down the line.

And my point is that I don't see good enough evidence that the effects involved are real to justify either of them.

Using speculative extrapolations from social psychology to promote social engineering is dangerous and proves too much. Of course, one should still be *nice*, and a big part of niceness is judicious exercise of the virtue of silence . But trying to institute and enforce said virtue on a social level requires subtlety that I have not yet seen anyone involved show the slightest sign of possessing.

Post Scriptum

Think quick! What is your brain's number one thought upon hearing "Comoros"?

Nature is Not a Slate. It's a Series of Levers.

Last week I criticized pop social psychology while maintaining that social psychology itself was a pretty interesting window into human thought processes.

I was then handed a link to someone who apparently likes social psychology a lot less than I do.

You should read the whole thing, but here are the parts I'll be talking about. As you can see, it's kind of a conservative perspective saying social psychology is a liberal enterprise to deny human nature in favor of people being infinitely malleable based on their situation:

> Personally, I find it very hard to fathom the idiocy of Mischel's conclusion. It would mean that a person who others think of as for instance shy is really nothing of the kind. It just looks that way because we have only had chance to observe him or her in situations that elicits shyness. And if you think of yourself as shy you must be either plain wrong or stuck in a series of situations that by coincidence predisposes you to acting shy. This idea may sound like a joke, but the zeitgeist of the 1960s was left of sanity and lots of "intellectuals" believed Mischel the way they believed in Marx, Lenin and Mao.
>
> For that reason, social psychology became a major branch of psychology. After all, if it was all in the situation then this was the important field of research. Personality barely survived and its proponents, like Hans Eysenck and Arthur Jensen, were often dismissed as racists and right-wing lobbyists.

Unsurprisingly, it soon became evident that Mischel was wrong – there really was such a thing as a shy person. As traits became real again, personality psychology grew but at the same time social psychologists kept a grip on their dominant position by introducing interactionism, the study of both situations and personality. This way they blurred the line between the fields and managed to claim a lot of the newly available positions in personality research. But at heart they were never interactionists; they started out as situationists because of their political views and they have stayed that way ever since. To this day they rarely perform experiments in which personality measures are used. Their focus is very much on the situation. Look at the collective social psychology blog in the links to the right of this post – it's even called The Situationist, not The Interactionist. For most of these psychologists, interactionism was just a word with which to neutralize the enemy.

And today? Well, it's like the French say, the more something changes, the more it stays the same. This can be seen in a recent post in the above mentioned The Situationist (which is still interesting to follow because not all social psychology is crap) about Harvard Professor Francesca Gino's book Sidetracked, in which she describes how small things or situations derail our plans, intentions and even our morals. As an example she mentions an experiment in which she and psychologist Dan Ariely equipped participants with high-end sunglasses. Half of the participants were told that they were actually counterfeits, while the other half were told they were the real deal. The participants were then instructed to perform a mathematical task which left

room for cheating. It turned out that 70 percent of those who thought they wore knock-off sunglasses cheated compared to 30 percent in the other group.

This may sound like compelling evidence for the power of the situation, but is it really? The participants were all young women rather than a representative sample. But more importantly, they were informed that they were participating in a psychological experiment and then told to wear counterfeit sunglasses. That's pretty far from any kind of real life situation. It's more like saying, "let's play a game – you will be the bad guy." It supports the idea that social psychology is, as someone put it, a list of how people behave in weird situations. Needless to say, Gino and Ariely didn't use any personality measure since that would only distract attention from the power of the almighty Situation.

So if wearing fake sunglasses can make a person dishonest, how about the situation of being brought up by criminal parents? Now that should be a way more powerful situation. Psychologist Sarnoff Mednick and colleagues investigated this in the mid 1980s using data from over 14 thousand nonfamilial adoptions (in which the adoptive parents are unrelated to the child). They found that when both biological and adoptive parents had no criminal convictions the adopted child was eventually convicted in 13.5 percent of the cases, so that's our baseline. When adoptive parents had convictions but biological parents had not, the number of convicted adoptees only rose very slightly to 14.7 percent. So fake sunglasses will have a profound effect on your honesty, but being brought up by criminals will only marginally

> elevate your risk of being convicted of a crime. That must be some sunglasses.
>
> […]
>
> [Social psychologists] just need to construct some even more artificial situations in order to deliver those results that will prove that Marx was right all along. And no outsiders need to concern themselves with exactly how they go about doing that.

Where to start, where to start?

First of all, it is too bad that Staffan finds the importance of situation hard to believe, but at least he is in good company. They do not call this problem the Fundamental Attribution Error because it is rare (or for that matter because it is a correct and tenable position). In experiments, people consistently overestimate the effect of personality and underestimate the effect of situation. Insofar as social psychologists are the people trying to correct that, they are doing God's work.

On the other hand, there are also personality psychologists, who study personality. They, too, are doing good work. Contrary to Staffan's assertions, they are an integral and well-beloved part of the field of psychology. Of the two personality psychologists Staffan claims were villainized and dismissed, both made the list of the fifty most eminent psychologists of the 20th century. Eysenck is the third most-cited psychologist of all time.

Jensen was indeed "dismissed as a racist and right-wing lobbyist", but this was less because he dared to study personality and more because he spent much of his time trying to prove black people were genetically less intelligent and

because he took large amounts of money from right-wing organizations. This seems like the sort of case where "racist and right-wing lobbyist" might be a perfectly acceptable value-neutral description – although I agree Jensen's work, much of which was completely unrelated and brilliant, hasn't gotten the recognition it deserves.

Because of the success of personality psychology, some of Staffan's claims about social psych are a No True Scotsman argument. Personality psychology is considered a different field than social psychology. When Staffan notes that social psychologists accept the importance of personality but don't study it, this is exactly symmetrical to personality psychologists accepting the importance of situation but not studying it. So getting upset at social psychologists for not studying personality is a lot like getting upset at cardiologists for not studying the liver.

The sunglasses experiment? Staffan correctly points out (as I did last week) that several similar experiments are under challenge, but those were on unconscious priming whereas the sunglasses experiment (from the little I read on it) seems to be about conscious priming, which is on firmer ground. To me it seems like exactly the sort of thing that might be true – in the extremely artificial conditions of the laboratory. This meshes with the point I made last week – that social psychology does a great job illuminating certain processes the brain goes through, but that we should be wary about assuming they have what doctors call "clinical significance".

The criminal adoption experiment? Probably 100% correct. As anyone who's read *The Nurture Assumption* (or my review thereof) knows, psychologists are *constantly* unable to find any effect of parents on their childrens' personalities or actions. This is sufficiently distressing that most people refuse

to believe it, but it keeps being confirmed again and again. Personality is 50% genetic and 50% some other factor which people have yet to illuminate but which definitely doesn't involve upbringing (I hear Judith Rich Harris' book *No Two Alike* purports to explain what this factor is, but I'm only halfway through it and can't comment).

(seriously, why do I have to spend so much time insisting to racists and eugenicists that *genes are seriously really important*? This should not be as big a niche in the blogosphere as it is!)

But basically, social psychology has discovered the correct fact that situation is more important than people think it is and personality is less important than people think it is in determining behavior, while not denying that both are pretty important. It correctly claims that priming can have very large short-term effects on unimportant decisions, and correctly notes that being raised by criminals has no effect on anyone's personality. So far I think it's doing pretty well, as long as, once again, you are skeptical about trying to do social engineering with its results outside the laboratory.

Now let's get to the part about Marx. For this we go to those experts on all things Marxist, More Right. In their latest post, Drew Summitt draws a distinction between what he considers a conservative view – that there is such a thing as human nature and that political systems need to take that into account – and a progressive view – that there is no such thing as human nature, people are infinitely malleable, and once we create some kind of utopia we can perfect mankind. The quote is his, the emphasis is mine:

> "He defends this proposition with the assertion that "the **conservative realist view of human imperfectability**

and their commitment to ordered liberty **as rooted in nature**, custom, and prudence“ can see great support in modern evolutionary theory because modern evolutionary theory contradicts what Thomas Sowell calls **the “unconstrained vision” that liberal intellectuals and theorists are tempted to hold**. In contrast to this Sowell sets up **a “constrained vision” of human anthropology that is limited in its capabilities by an intellect being the servant of the passions, the reality of sin, or boring genes telling our memes what they can and can’t do**. In order to support the idea that evolutionary theory supports a conservative political vision Arnhart traces the foundations of human capability to nature, custom and prudence. The conservative hierarchy of nature, custom and prudence **is what constrains the idealist impulse for reason to govern and judge, and indeed seek to overthrow, custom and nature**

[…]

But the Liberal Egalitarian Free-trade Technocrats also recognize human nature as being essential to political order. If you agree with Steven Pinker on the Humean Is in regard to human nature, that is, if you think he’s got his facts right, you must have an independent account of the Humean Ought, because Pinker is a Liberal, though one of the deflated, Clintonian Liberals. This is a form of Liberalism not touched enough reactionary circles. Why is it the Daniel Dennett, Steven Pinker, Peter Singer and many other experts in contemporary studies in human nature self described liberals? They may reject modern queer theory, as Dawkins does, or they may think that the brain is not a Blank Slate, as Pinker does, but they don’t

> consider these positions to be dangerous to Liberalism writ large or if they do they are terribly good at hiding it.

So he is wondering why, if liberalism is founded on the idea that human nature is an infinitely malleable blank slate, are the world's greatest scientific experts in evolved human nature liberals?

To steal a delightful turn of phrase from Terry Eagleton, this is like wondering why, if Tony Blair is an octopus, he has only two arms.

Let me make an analogy to medicine. Unlike the brain, there is no debate on the "nature" of the heart – the literal blood-pumping heart, not the fuzzy emotional version. We know the heart is fully one hundred percent genetically programmed (minus a little morphogenetic variation), that it's not malleable by schooling or brainwashing or being raised in a commune. It is a social engineer's nightmare, a system founded entirely on human nature without the slightest wiggle room.

And yet doctors *routinely* make the heart do what they want. If they want to raise heart rate, they give a dose of epinephrine. If they want to lower heart rate, they give a dose of propranolol. If social planners could control the brain as easily as doctors control the heart, we'd already be living in a communist utopia.

The heart has an immutable nature, and that immutable nature is *to respond to different situations in highly predictable ways*.

The heart is neither a blank slate nor a fully inscribed slate. It's not a slate at all. The heart is a series of levers. If you pull one lever, it will do one thing. If you pull another lever, it will do another thing. It is, paradoxically, hard-coded for malleability. It's not infinitely malleable – there's no drug you can inject to

make the patient's heart beat out the drum parts to Beatles songs – but you can shift it a little bit this way or that.

We have reason to believe the brain works the same way. Not everything is a lever – if you send a kid off to be raised by criminals, it won't activate any of the hard-coded IF-THEN statements, and nothing will happen. But if you surround someone by stimuli that prime the idea of criminality – whether sunglasses or a broken window, that will pull on a lever that will make criminal behavior a little bit more likely.

(except for "levers", read "extremely complicated things that run through chaos theory at some point and so are inherently unpredictable except in the broadest and most statistical sense")

All of this reminds me of a video I saw this afternoon on the second day of The Hospital Orientation. Please excuse me if I change it around just a little to turn it from a quality improvement case study to a morality tale.

There were two hospitals, Hospital A and Hospital B. Both, like all hospitals, were fighting a constant battle against medical errors – surgeons removing the wrong leg, doctors giving the wrong dose of medication, sleepy interns reading x-rays backwards, that kind of thing. These are deadly – they kill up to a hundred thousand people a year – and terrifyingly common.

Hospital A took a very right-wing approach to the issue. They got all their doctors together and told them that any doctor who made a minor medical error would get written up and any doctor who made a major medical error would be fired. Rah personal responsibility!

Unfortunately, when they evaluated the results of their policy they found they had exactly as many medical errors as before,

except now people were trying to cover them up and they weren't being discovered until way too late.

Hospital B took a very progressive approach. They too got all their doctors together, but this time the hospital administrators announced: "You are not to blame for any medical errors. If medical errors occur, it means we, the administrators, have failed you by not creating a sufficiently good system. Please tell us if you commit any medical errors, and you won't be punished, but we will scrutinize what we're doing to see if we can make improvements."

Then they made sweeping changes to what you might call the "society" of the hospital. They decreased doctor workload so physicians weren't as harried. They shortened shifts to make sure everyone got at least eight hours of sleep a night. They switched from paper charts (where doctors write orders in notoriously hard-to-read handwriting) to electronic charts (where everything is typed up). They required everyone to draw up and use checklists. They even put propaganda posters over every sink reading "DID YOU WASH YOUR HANDS LONG ENOUGH??!" with a picture of a big eye on them. You can't get more Orwellian than that.

And yet, *mirabile dictu*, this was the hospital that saw their medical error rates plummet.

The administrators of this second hospital didn't ignore human nature. Instead, they exploited their knowledge of human nature to the fullest. They know it's in human nature to do a bad job when you're working on no sleep. They know it's human nature to try to cut corners, but that people will run through checklists honestly and effectively. They even know that studies show that pictures of eyes make people behave more prosocially because they feel like they're being watched.

You don't have to tell me all the reasons this doesn't directly apply to an entire country. I can think of most of them. But my point is that if I'm progressive – a label I am not entirely comfortable with but which people keep trying to pin on me – this is my progressivism. The idea of using knowledge of human nature to create a structure with a few clever little lever taps that encourage people to perform in effective and prosocial ways. It's a lot less ambitious than "LET'S TOTALLY REMAKE EVERY ASPECT OF SOCIETY AS A UTOPIA", but it's a lot more practical.

(Although I'm also kinda okay with making every aspect of society a utopia, as long as we do it right.)

The Anti-Reactionary FAQ

[Edit 3/2014: I no longer endorse all the statements in this document. I think many of the conclusions are still correct, but especially section 1 is weaker than it should be, and many reactionaries complain I am pigeonholing all of them as agreeing with Michael Anissimov, which they do not; this complaint seems reasonable. This document needs extensive revision to stay fair and correct, but such revision is currently lower priority than other major projects. Until then, I apologize for any inaccuracies or misrepresentations.]

0: What is this FAQ?

This is the Anti-Reactionary FAQ. It is meant to rebut some common beliefs held by the political movement called Reaction or Neoreaction.

0.1: What *are* the common beliefs of the political movement called Reaction or Neoreaction?

Neoreaction is a political ideology supporting a return to traditional ideas of government and society, especially traditional monarchy and an ethno-nationalist state. It sees itself opposed to modern ideas like democracy, human rights, multiculturalism, and secularism. I tried to give a more complete summary of its beliefs in Reactionary Philosophy In An Enormous, Planet Sized Nutshell.

0.1.1: Will this FAQ be a rebuttal the arguments in that summary?

Some but not all. I worry I may have done too good a job of steelmanning Reactionary positions in that post, emphasizing what I thought were strong arguments, sometimes even correct arguments, but not really the arguments Reactionaries believed or considered most important.

In this FAQ, I will be attacking not steel men but what as far as I can tell are actual Reactionary positions. Some of them seem really dumb to me and I excluded them from the previous piece, but they make it in here. Other points from the previous post *are* real Reactionary beliefs and make it in here as well.

0.2: Do all Reactionaries believe the same things?

Obviously not. In particular, the movement seems to be divided between those who want a feudal/aristocratic monarchy, those who want an absolute monarchy, and those who want some form of state-as-corporation. Even more confusingly, sometimes the same people seem to switch among the three without giving any indication they are aware that they are doing so. In particular the difference between feudal monarchies and divine-right-of-kings monarchies seems to be sort of lost on many of them.

In general, this FAQ chooses two Reactionary bloggers as its foils – Mencius Moldbug of Unqualified Reservations, and Michael Anissimov of More Right. Mencius is probably the most famous Reactionary, one of the founders of the movement, and an exceptionally far-thinking and knowledgeable writer. Michael is also quite smart, very prolific, and best of all for my purposes unusually willing to state Reactionary theories plainly and explicitly in so many words and detail the evidence that he thinks supports them.

Mencius usually supports a state-as-corporation model and Michael seems to be more to the feudal monarchy side, with both occasionally paying lip service to divine-right-of-kings absolutism as well. Part 2 of this FAQ mostly draws from Michael's feudal perspective and Part 4 is entirely based on Moldbug's corporation-based ideas.

0.3: Are you going to treat Reaction and Progressivism as real things?

Grudgingly, yes.

One of the problems in exercises like this is how much to take political labels seriously. Both "Reaction" and "Progressivism" are vast umbrella concepts on whose definition no one can agree. Both combine many very diverse ideas, and sometimes exactly who falls on what side will be exactly the point at issue.

Part of Part 3 will be an attempt to define Progressivism, but for now I'm going to just sweep all of this under the rug and pretend that "Reactionary" and "Progressive" (or for that matter "leftist" and "rightist") have obvious well-defined meanings that are exactly what you think they are.

The one point where this becomes very important is in the discussion over the word "demotist" in Part 2. Although debating the meaning of category words is almost never productive, I feel like in that case I have *more* than enough excuse.

1: Is everything getting worse?

It is a staple of Reactionary thought that everything is getting gradually worse. As traditional ideas cede to their Progressive replacements, the fabric of society tears apart on measurable ways. Michael Anissimov writes:

> The present system has every incentive to portray itself as superior to all past systems. Reactionaries point out this is not the case, and actually see present society in a state of severe decline, pointing to historically high levels of crime, suicide, government and household debt, increasing time preference, and low levels of civic participation and self-reported happiness as a few examples of a current cultural and historical crisis.

Reactionaries usually avoid getting this specific, and with good reason. Now that Michael has revealed the domains in which he is critiquing modern society, we can start to double-check them to see whether Progressivism has indeed sent everything to Hell in a handbasket.

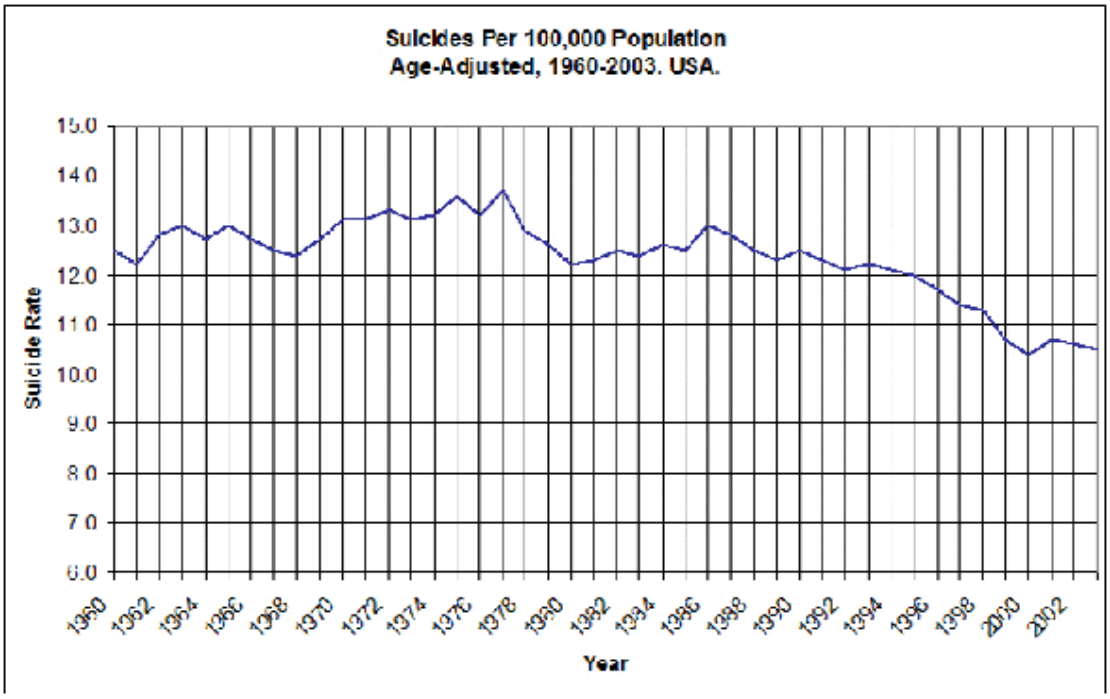
But I must set some strict standards here. To support the Reactionary thesis, I will want to see long-term and unmistakeable negative trends in these indicators. Nearly all Reactionaries agree that the advance of Progressivism has been a long-term affair, going on since the French Revolution if not before. If the Reactionaries can muster some data saying that something has been getting better up until 2005 but declining from 2005 to the present, that doesn't cut it. If something else was worsening from 1950 to 1980 but has been improving since then, that doesn't cut it either. I will not require a completely monotonic downward trend, but neither will I accept a blip of one or two years in a generally positive trend as proving all modern civilization is bankrupt.

Likewise, if something has been getting worse in Britain but not the United States, or vice versa, that will not suffice either. Progressivism is supposed to be a worldwide movement, stronger than the vagaries of local politics. I will not require complete concordance between all Western countries, but if the Anglosphere countries, France, Germany, and Japan seem split about fifty-fifty between growth and decay in a certain indicator, blaming Progressivism isn't going to cut it.

So, without further ado, let's start where Michael starts: with suicide.

1.1: Is suicide becoming more common?

Here's the US suicide rate from 1960 to 2002:



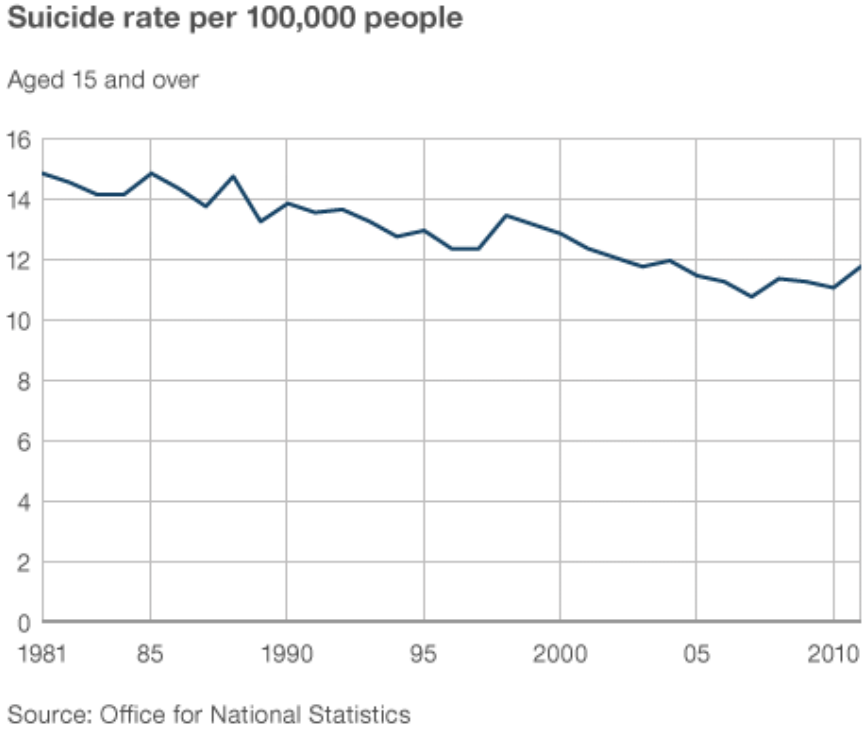
(Source: US Census Bureau. Statistical Abstracts of the United States. 2004-2005. "No. 102: Deaths by Major Causes 1960-2002 (Age Adjusted Death Rates Per 100,000 Population.)" www.census.gov/prod/2004pubs/04statab/vitstat.pdf; Center for Disease Control and Prevention, National Center for Health Statistics. National Vital Statistics Report, vol. 53, no. 15, Feb. 8, 2005, "Deaths: Preliminary Data for 2003" Table 2, "Deaths, Death Rates, and Age Adjusted Death Rates, for 113 Selected Causes, Injury by Firearms, Drug-Induced Deaths, Alcohol Induced Deaths, and Injury at Work: United States Final 2002 and Preliminary 2003." Online. www.cdc.gov/nchs/data/nvsr/nvsr53/nvsr53_15.pdf)

In those forty years, considered by many the heyday of the leftist movement, forty years encompassing the Great Society, the civil rights movement, the explosion of feminism onto the public consciousness, the decline of the traditional family, etc, etc… suicide rates dropped about 20%.

What evidence have the Reactionaries cite for their side? Michael cites a New York Times article pointing out that suicide rates rose from 1999 to 2010. Apparently my new job is reminding Reactionaries that they cannot blindly trust New York Times articles to give them the whole truth.

Suicide rates did rise from 1999 to 2010. But if we're going to blame leftism for rising suicide rates it's kind of weird that it would choose the decade we had a Republican President, House, Senate, and Supreme Court to start increasing. A more likely scenario is that it had something to do with the GIANT NEVER-ENDING RECESSION going on at the time.

As we mentioned above, since Reactionaries believe that Progressivism has been advancing simultaneously in many different countries it is worthwhile to check whether other nations show the same trends as the United States. If every country that was becoming more Progressive showed increased suicide rates, this would be strong evidence that Progressivism were to blame. But if some Progressive countries experienced lower suicide rates, that would suggest country-specific problems.



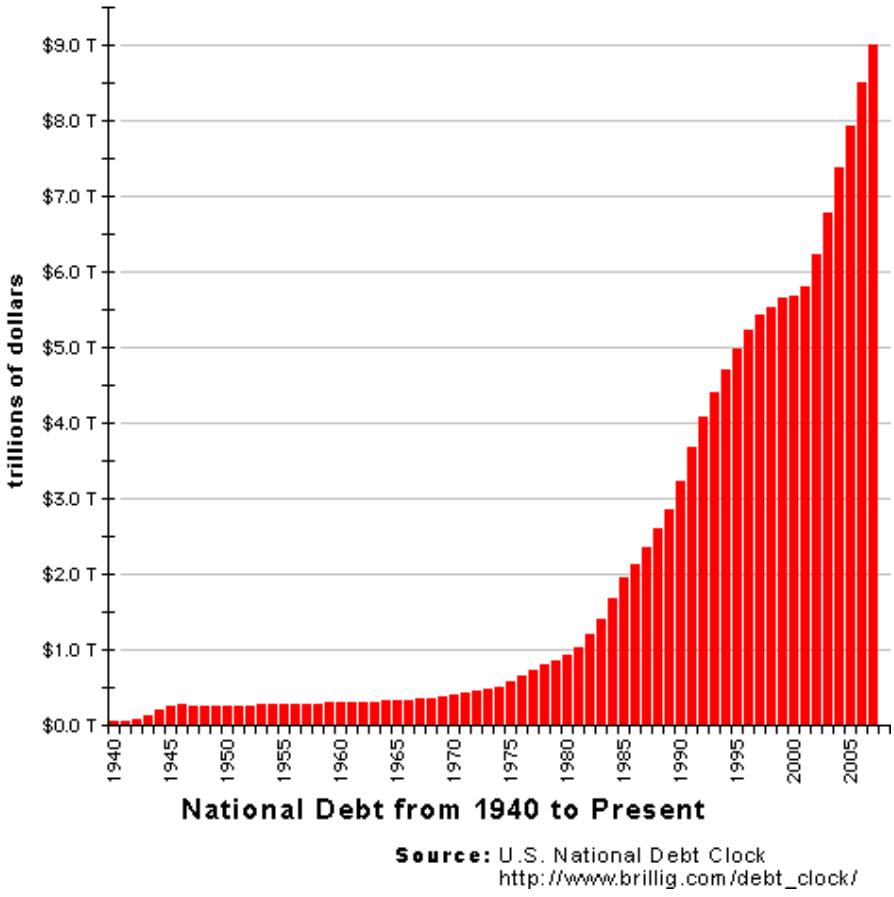
In Britain, we find not only that suicide has generally been going down for the past thirty years, but that – as predicted above – there is a bit of an upward tick corresponding with the Great Recession.

Even better, we find that suicide peaked in Britain in 1905 – just after the Victorian period – and has been declining ever since.

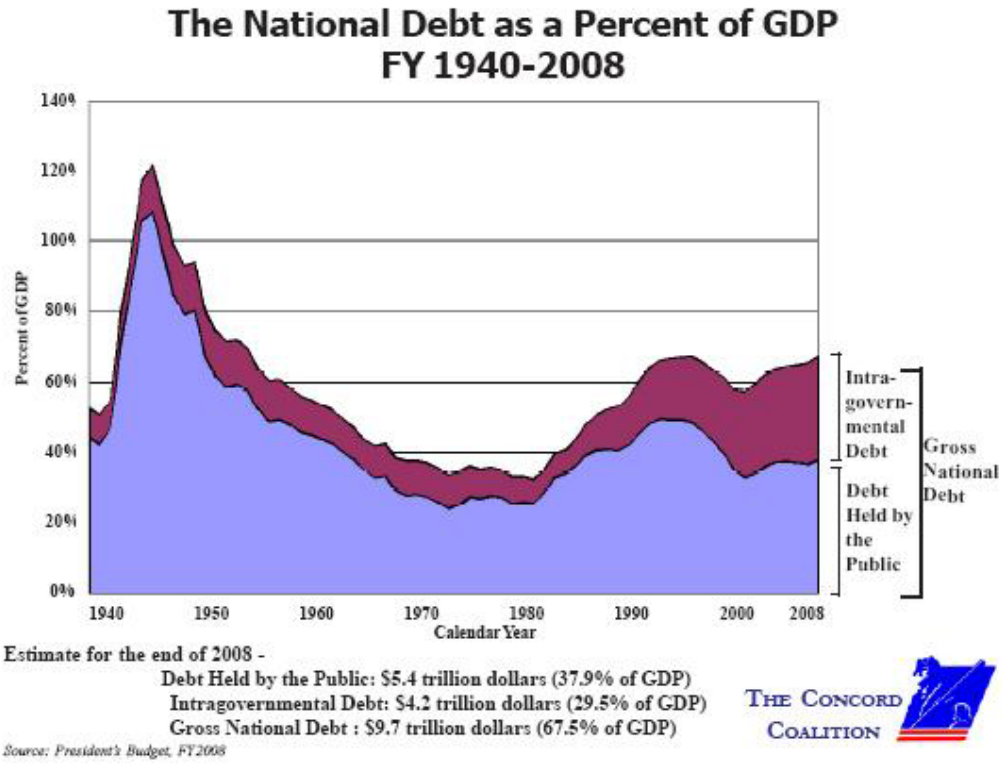
I try to be nice. I really do. But I will say it – the Reactionary argument that suicide has been increasing during modernity from a low during some fantasized Victorian Golden Age is *unacceptably shoddy*.

1.2: Is everyone falling further and further into debt?

Here again the Reactionaries overstate their case. Michael tried to support his point with…



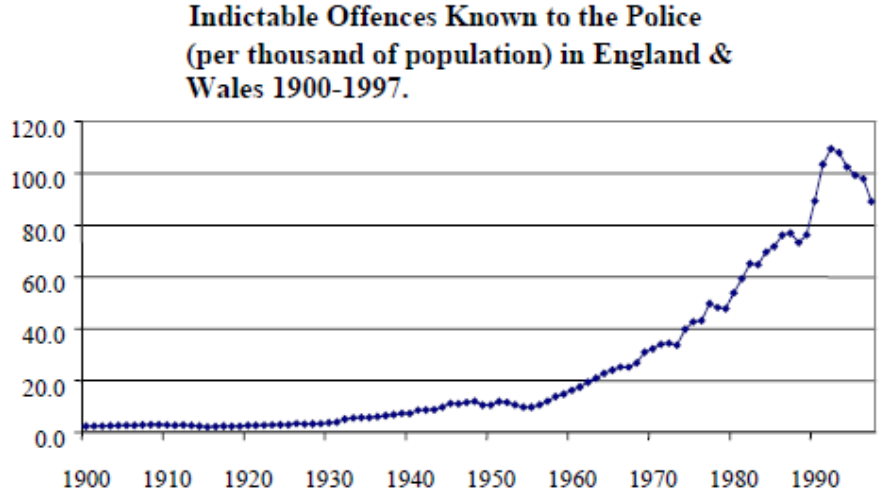
…which shows government debt rising ceaselessly and alarmingly through the simple tricks of not adjusting for inflation or rising GDP. Keep yourself honest by taking those steps, and the situation looks more like this:



To his credit, Michael fixed this when I pointed it out. But to me, the new graph looks like gradual decrease in debt since World War II up until Reagan's big military buildup, followed by a gradual retreat from that military buildup. My God, won't somebody stop Progressivism before it's too late?!?!

1.3: Is crime becoming worse?

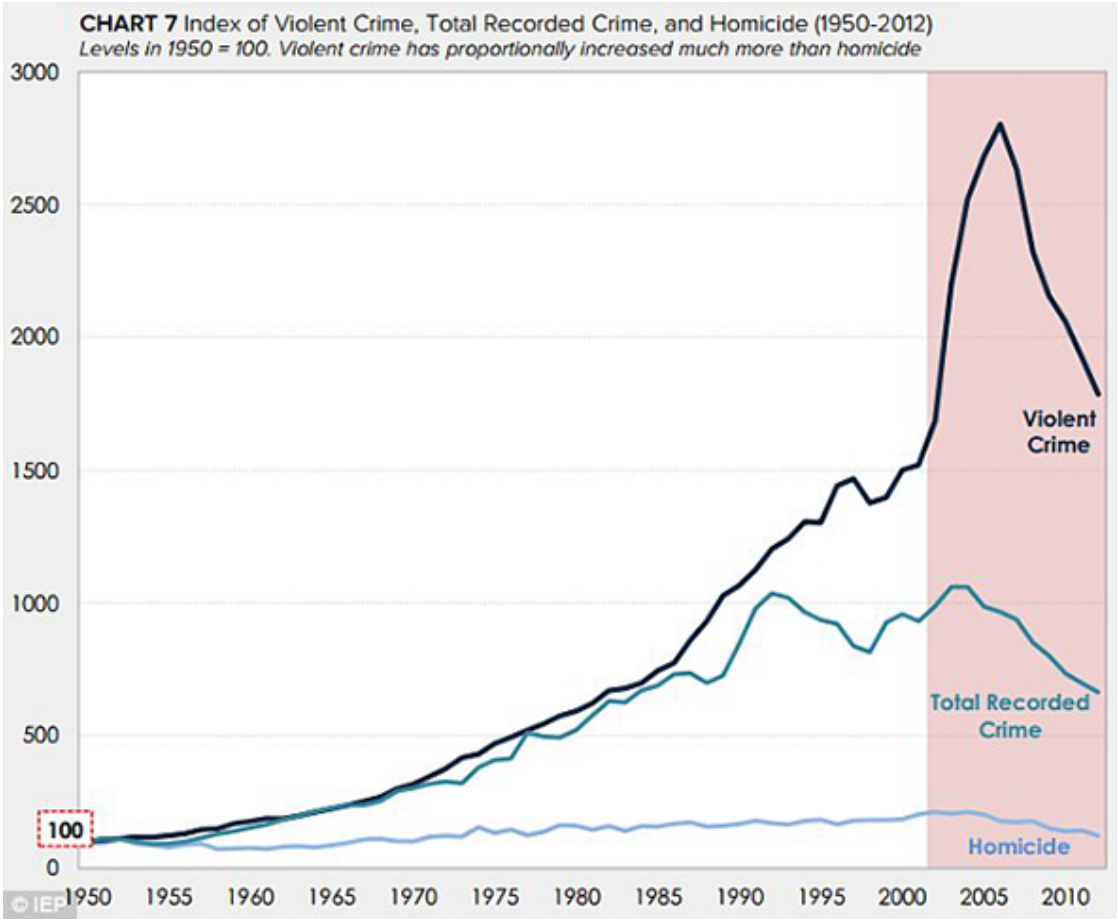
Michael's statistics for crime deserve more attention:



Question number one: what does this graph mean by “indictable offenses”? This very broad term introduces no fewer than three dangerous biases. First, we have reporting bias – the more police there are and the more active there are, the more crimes get heard about and reported. Second, we have definition bias within individual crimes – for example, larceny in Britain fell by two thirds in 1855, but this was because Parliament passed a law raising the minimum amount of property that had to be larcened for it to count. Third, we have broader definition bias in what is or isn’t a crime – how much of that rise around 1970 was the “indictable offense” of people smoking marijuana, something that was previously neither illegal nor widely available?

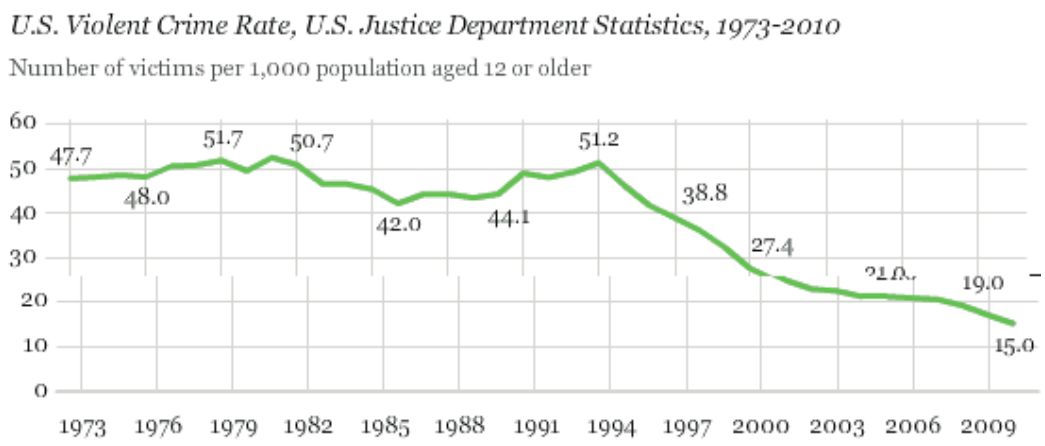
Criminologists’ recommended way around this problem is to look at *murder*. The murder rate tends to track the crime rate in general. Murder isn’t as subject to reporting bias – if someone is killed, the police are going to want to hear about it no matter how understaffed they are. And murder is less subject to changes in definition – dead is dead.

So let’s add the homicide rate to the above chart:

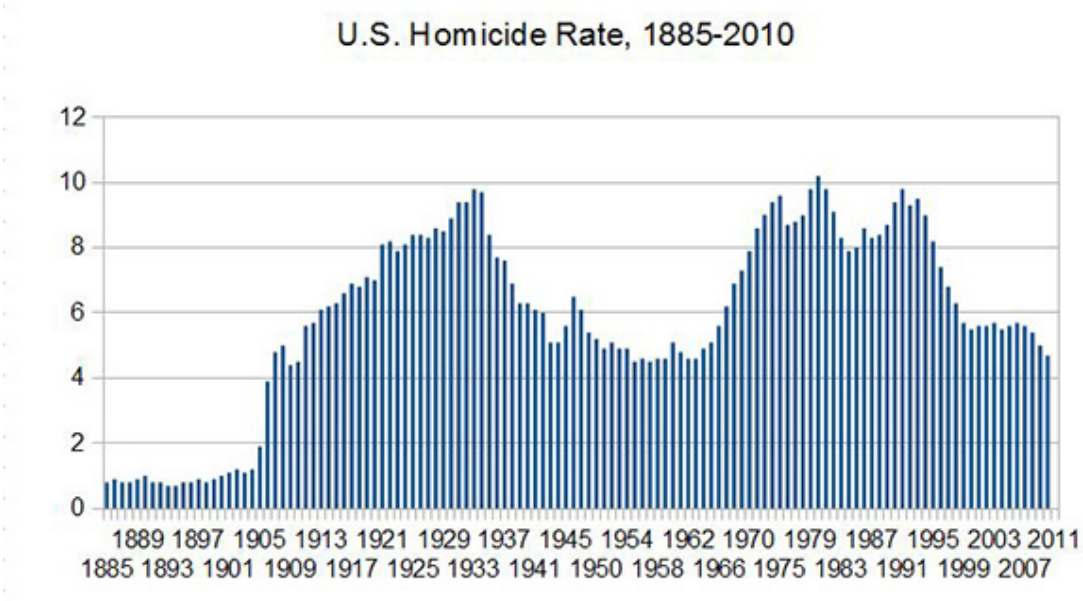


Alas, I can only find the numbers since 1950 rather than 1900. But as we can see, despite the huge rise in “violent crime”, homicide rates stay very steady and perhaps even decline a little over that period.

Question number two: Michael is American. All his other statistics make reference to American numbers. Why does he suddenly switch to Britain when we talk about crime? I won’t impugn his motives – long-term US crime data is really hard to find. But it’s worth pointing out that what there is, is much less sensational:



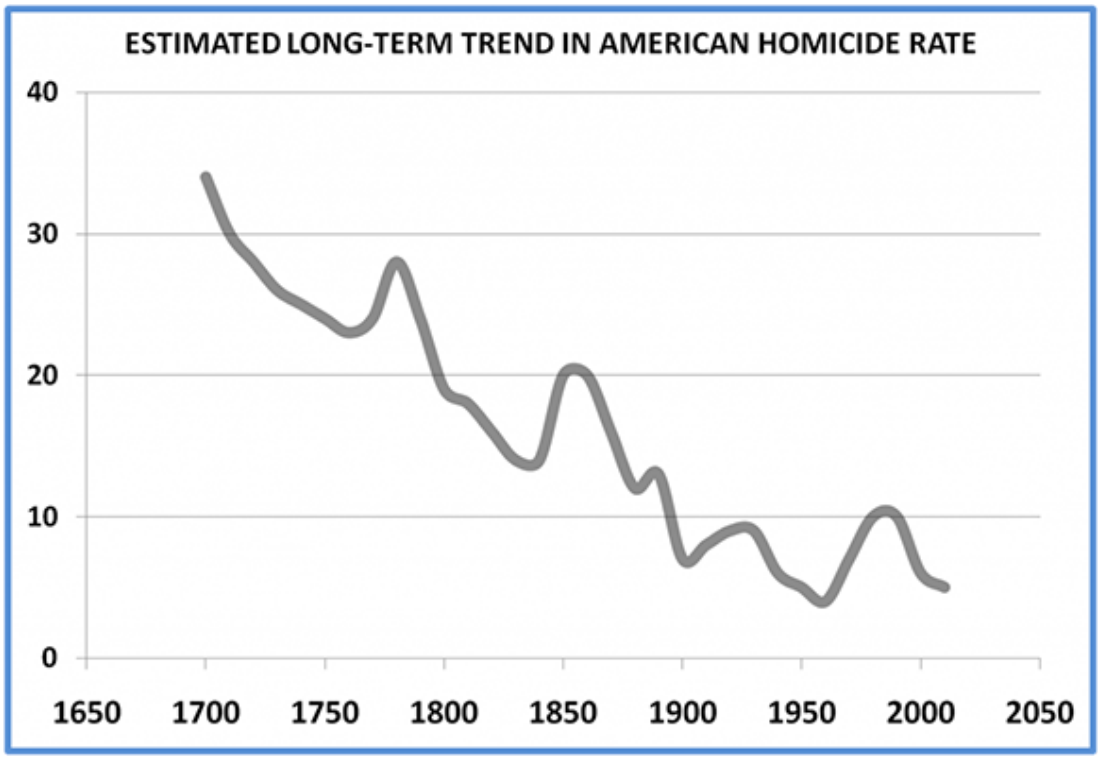
I wish I could find longer-term US crime rate data, but it doesn't seem to be out there. I can, however, find longer-term homicide data:



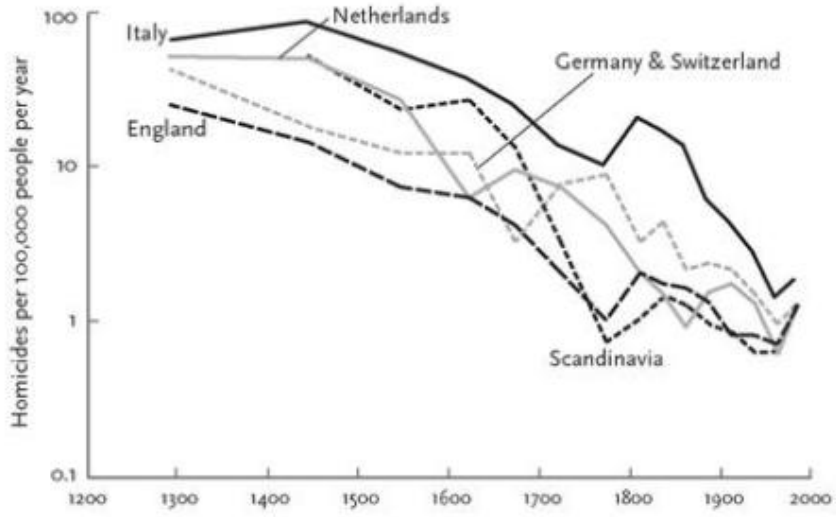
We see ups and downs but no general pattern. A Reactionary might cite the apparently very low level of homicides in 1885, but historians pretty much agree that's a reporting artifact and that the period ending in 1887 had the highest murder rate in American history. In any case, right now we seem to be enjoying a 50 year low. And lest someone bring up that medical technology has advanced enough to turn many would-be murders into attempted murders – which is true – aggravated assaults, the category of crime that would encompass attempted murders, are less than half

of what they were twenty years ago. Kind of hard to square with everything getting worse and more violent all the time.

Actually, stopping at 1885 is for losers. Let's go *really* long-term. From Marginal Revolution, themselves drawing from Manuel Eisner's Long-Term Historical Trends in Violent Crime:



We've got to go deeper! From HBD Chick, citing Steven Pinker:



1.3.1: But the Victorian Era had amazingly low crime rates! People could walk out in any corner of the country unmolested!

Crime was basically a half-forgotten memory!

This is one of Mencius Moldbug's favorite points. He cites approvingly an 1870s British text which says that

> Meanwhile, it may with little fear of contradiction be asserted that there never was, in any nation of which we have a history, a time in which life and property were so secure as they are at present in England. The sense of security is almost everywhere diffused, in town and country alike, and it is in marked contrast to the sense of insecurity which prevailed even at the beginning of the present century. There are, of course, in most great cities, some quarters of evil repute in which assault and robbery are now and again committed. There is perhaps to be found a lingering and flickering tradition of the old sanctuaries and similar resorts. But any man of average stature and strength may wander about on foot and alone, at any hour of the day or the night, through the greatest of all cities and its suburbs, along the high roads, and through unfrequented country lanes, and never have so much as the thought of danger thrust upon him, unless he goes out of his way to court it.

Reactionaries take this idea and run with it – past societies were so well-organized that they had completely eliminated crime, whereas our own democratic government turns a blind eye while thousands of people are beaten and mugged and murdered and…

Again, let's concentrate on "murdered". It's the only crime that gives us a shot at apples-to-apples comparison. So what was the Victorian murder rate?

> Homicide is regarded as a most serious offence and it is probably reported more than other forms of crime. Between 1857 and 1890, there were rarely more than 400 homicides reported to the police each year, and during the 1890s the average was below 350. In Victorian England, the homicide

> rate reached 2 per 100,000 of the population only once, in 1865. Generally, it was about 1.5 per 100,000 falling to rarely more than 1 per 100,000 at the end of the 1880s and declining even further after 1900. These figures do not take into account the significant number of infanticides that went undetected. The statistics for homicide are therefore probably closer to the real level of the offence.

So, Victorian murder rate of between 1 and 2 per 100,000 people. And the current British murder rate? According to the United Nations Office on Drugs and Crime, it stands at 1.2 per 100,000 people, rather lower than the Victorian average.

1.3.1.1: But if the Victorian crime rate was as high or higher than it is today, how come Victorians felt completely safe and thought that crime had been eradicated?

Normally this is where I'd start talking about how we moderns are constantly exposed to so many outrageous and terrifying stories in the media that we don't realize how good we have it. But in this case that turns out to be explaining away a nonproblem. The Victorians were absolutely terrified of crime and thought they were in the middle of a gigantic crime wave. Here's Understanding The Victorians on the "garroting panic":

> Violent attacks by strangers were seen as grave cause for concern. There was a disproportionate amount of attention paid to violent nighttime assaults by strangers in urban areas, called "garroting" and similar to what we might call "mugging". There were garroting panics in 1856 and 1862, in part because of extensive press coverage. In the highest profile case, MP Hugh Pilkington as attacked and robbed in London at one o'clock in the morning on July 17, 1862, after leaving a late session in the House of Commons. Press reports of garroting increased dramatically, and the public quickly became convinced there was a serious problem. Garroting panic was so rampant that it became a topic of satire: *Punch*

> published several cartoons of men running from their own shadows or from trees that they were convinced were garrotters.

And A History of Criminal Justice In England and Wales on the same topic:

> Crimes of violence were perceived to be on the increase in the 1850s and panic set in when an outbreak of garrotting occurred in various parts of the country in the period from 1856 to 1862. Garrotting involved choking, suffocating, or strangling a victim. During these years, Punch magazine carried a whole series of cartoons and lengthy jokes about the crime, including many eccentric means of defense. One advertisement appeared offering the public an "anti-garrot collar". This was a steel collar to be hand-fitted round the neck with a large number of sharp steel spikes pointing outwards. Despite such bizarre forms of protection, the offence caused a great deal of fear among the public and it was generally regarded as a very serious threat to law and order. Letters to *The Times* began to appear from gentlemen who had been so attacked and robbed. In response the judges began to order severe floggings in addition to penal servitude in an attempt to stem the growth of the crime. Their example was then followed by Parliament which, against the wishes of the government, enacted the Security From Violence Act 1863.

So if there was so much panic about crime, how come the person who wrote Moldbug's favorite book felt Victorian Britain was crimeless?

I guess it all depends on your perspective. I live less than two miles outside Detroit city limits, and I've never been the victim of a single crime in my life or even felt particularly threatened. Some people just live sheltered existences.

But apparently most other Americans agree with me. According to Gallup, 89% of American men currently feel safe walking alone at night in the city where they live. If 89% of modern US men feel that way, I'm not surprised Moldbug could find one Victorian guy willing to express the opinion.

1.3.2: Why does this matter again?

For some reason, the Reactionaries have made crime an absolute linchpin of their case. A very large portion of Reactionary thought goes implicitly or explicitly through the argument "Progressives have legitimized minorities, minorities cause crime, crime is destroying our society, therefore Progressivism must be destroyed."

The extent of the Reactionary obsession with crime never fails to amaze me. Moldbug writes:

> Security and liberty do not conflict. Security always wins. As Robert Peel put it, the absence of crime and disorder is the test of public safety, and in anything like the modern state the risk of private infringement on private liberties far exceeds the official of public infringement. No cop ever stole my bicycle.

Desperate times call for desperate measures. On the other hand, non-desperate times call for non-desperate measures. And this is a time when everything is pretty much okay. Murder and violent crime are at historic lows, and almost 90% of American men feel safe walking outside at night. Crime is very nearly a non-issue, and when designing a system of government it is probably a bad idea to give them a blank check to ruin everything else in the pursuit of decreasing it.

1.4: Are people becoming less happy?

Michael's source for decreasing happiness levels is Blanchflower & Oswald: Well Being Over Time In Britain And The USA. But read the abstract, and you find it's more complicated: "Reported levels of well-being have declined over the last quarter of a century

in the US; life satisfaction has run approximately flat through time in Britain."

Once again, we find these supposed effects of a global trend are very much limited to individual countries.

Second, when we check the breakdown, we find, as the paper puts it, that "[American] men's happiness has an upward trend, yet American women's well-being has fallen through the years." At a guess, I'd say this is because more women are working full-time jobs. This may be a bit of a victory for Reactionaries, who are no fans of feminism, but it is a very limited victory with little broader implication for other aspects of society. If you're a man, there's never been a happier time to be alive.

Further, Blanchflower and Oswald aren't the only people trying to measure happiness. Ruut Veenhoven has collected 3,651 different happiness studies into a World Database of Happiness. Inglehart, Foa, and Welzel have sorted through some of the data and find that:

> Among the countries for which we have long-term data, 19 of the 26 countries show rising happiness levels. In several of these countries – India, Ireland, Mexico, Puerto Rico, and South Korea – there are *steeply* rising trends. The other countries with rising trends are Argentina, Canada, China, Denmark, Finland, France, Italy, Japan, Luxembourg, the Netherlands, Poland, South Africa, Spain, and Sweden. Three countries, the US, Switzerland, and Norway, show flat trends. Only four countries, Austria, Belgium, UK, and West Germany, show downward trends.

Investigating further:

> By far the most extensive and detailed time series comes from the US, and the full series covering the 60 years from 1946 to 2006 shows a flat trend. But the subset from 1946 to 1980 show a downward trend, while the series from 1980 to 2006 shows a rising trend. A similar picture appears from the much

> scantier British dataset. The entire series from 1946 to 2006 shows a downward trend, but the series from 1980 to the present shows a clear upward trend.

So there you have it. In 19/26 countries, happiness has risen since 1946, and in both America and Britain, it's been rising since 1980.

1.5: Is time preference decreasing?

Time preference is a mathematical formalization of whether people live only for the moment like the proverbial grasshopper, or build for the future like the proverbial ant. We'd probably prefer if people had pretty low time discounting (ie are more ant-like). Michael claims that in fact we're becoming more grasshopper-like.

He cites as his source Wang, Rieger and Hans' How Time Preferences Differ, which is a fascinating study but which does not, as far as I can tell, make anything like the claim Michael says it does. It seems to be entirely about comparing different countries. There is only one thing that looks even close to an intertemporal comparison:

> In particular, 68% of our [2011] US sample chose to wait. For comparison, in the survey by Frederick (2005) where he used the same question…only around 41% of students chose to wait.

Here we see people saving *more* over time, ie becoming more ant-like, although it would be absurd to think this represented a real effect over such a small time period.

Michael may be referring to a claim buried in the study that collectivism is linked to lower discount rates than individualism. This study was done entirely on Israeli Arabs and Jews, with Jews as a proxy for "individualist cultures" and Arabs as a proxy for "collectivist cultures". Suffice it to say this is not how broad human universals are established. A similar experiment compared Western-primed Singaporeans with Eastern-primed Singaporeans

to “conclude” that Confucian cultures had a “longer-term outlook” and thus a lower discount rate. This would be all nice and well except that in the main study, Canadians had a lower discount rate than Japanese, Chinese, Taiwanese, or Koreans. So much for Confucians.

1.6: Is civic participation decreasing?

The argument is simple. Democracy fractures traditionalist societies, destroying civic cohesion, which in turn reduces voter turnout. Therefore, the only way to increase voter turnout is to abolish democracy.

No, actually the argument is more complex, and Michael cites Robert Putnam’s Bowling Alone to make his point for him. Since there is no one statistic for civic participation, I can’t refute it with pure data the same as I tried to do with the others.

But I will point out that Putnam’s own thesis is that it is technology – our options of watching TV, playing video games, or hanging out on the computer – that make us less involved in our communities. He may be right. But blaming the politically neutral force of technology acquits Progressivism.

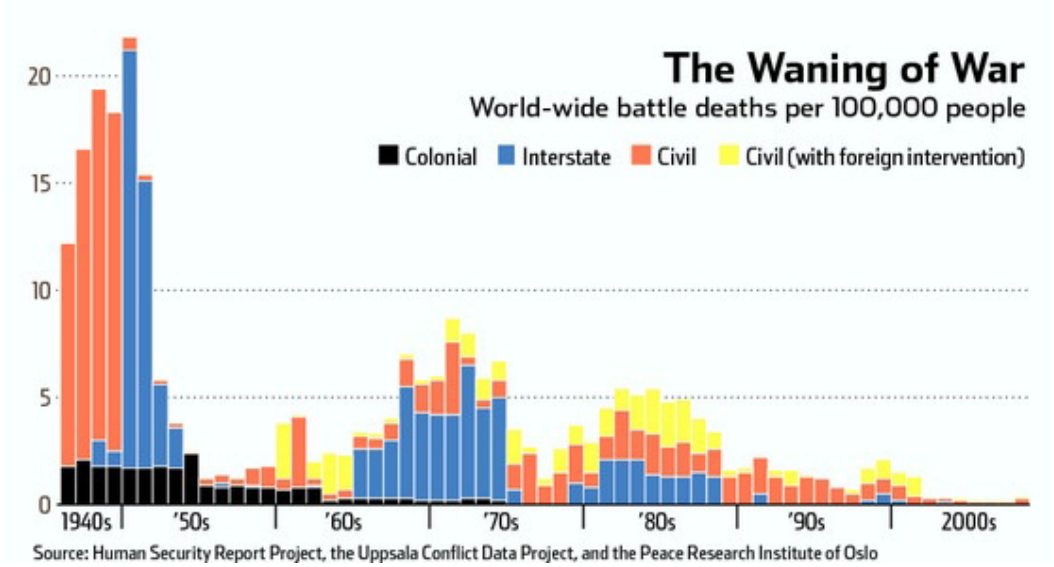
Even so, a word to defend technology. Right now I am typing a lengthy essay that will be read by a few thousand people. A couple dozen of those will discuss it in the comments. Among those will be people with whom I’ve had interesting discussions, friendships, and even a couple of romantic relationships. Through the ensuing debate, I will meet new people with whom I will likely keep in touch and discuss my extremely niche interests with on a near daily basis for many years to come, forming bizarre but intellectually fecund communities that will inevitably end up with everyone involved moving to the Bay Area and having kids together.

And we are supposed to be upset because the technology that makes this possible has *cut down on the number of bowling leagues*? That’s like condemning butterfly metamorphosis for decreasing the number of caterpillars.

1.7: Are international conflicts becoming more frequent?

This isn't in the paragraph quoted above, but Michael has expressed the opinion to me in person, and anyone familiar with Reactionary thought will recognize this as a staple. The theory is that monarchies had strong international law between them that prevented or settled conflicts quickly, but that democracies have the "sham" international law of the UN (exactly what makes it a sham is never explained) and constantly interfere in one anothers' business as a continuation of their own internal politics or obsession with human rights.

As far as I know no Reactionary has ever dared to cite statistics that they say support this claim, which is probably for the better. But just for the record, here's the counterclaim:



You can find a much more exhaustive discussion of this topic here.

1.7.1: What about the Concert of Europe? The great statesman Klemens von Metternich used Reactionary ideas to create a brilliant system that kept peace in Europe for nearly a century!

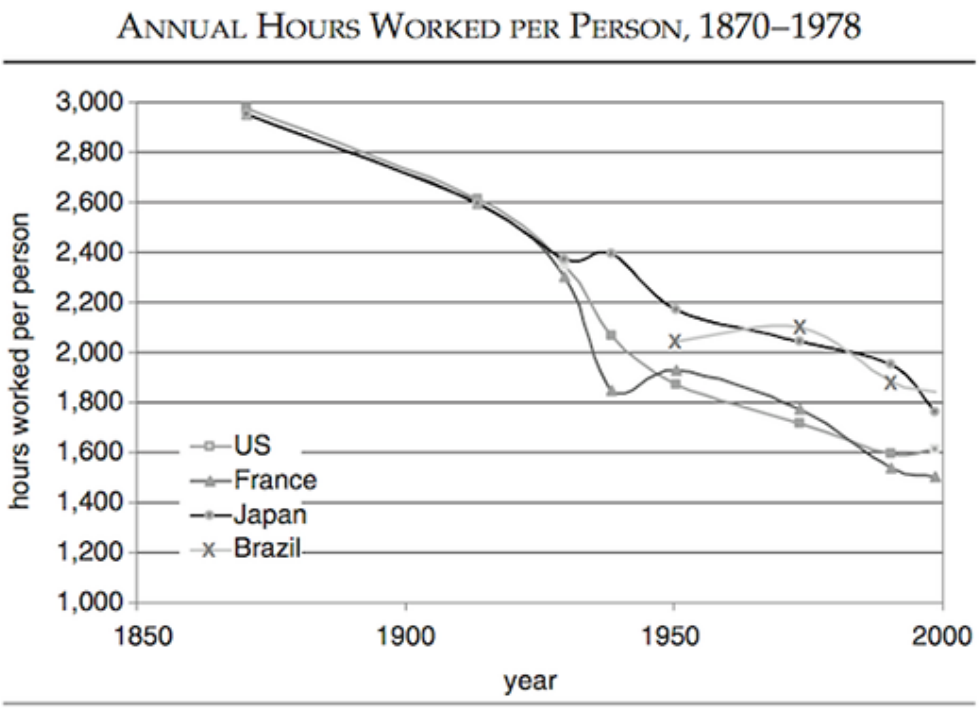
The Concert of Europe lasted from 1815 to 1914. During that time, Europe suffered – just counting major interstate wars involving Congress of Vienna participants – the French Invasion of Spain, the Crimean War, the Schleswig Wars, the Wars of Italian

Independence, Austro-Prussian Wars, the Franco-Prussian War, and, let's not forget, World War I.

The modern equivalent of the Concert of Europe is the European Union, but built on Progressive rather than Reactionary principles. It has existed from 1951 to 2013 so far, and In those sixty-two years, major interstate wars between EU members have included… well, none.

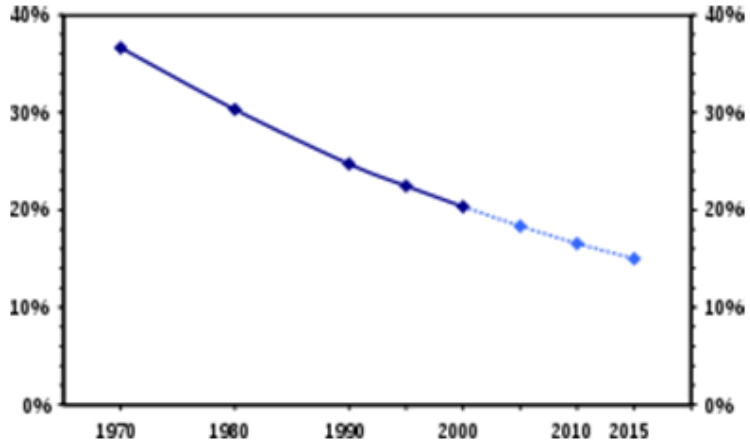
1.8: Okay, you've discussed the trends Michael listed as supporting Reaction, and found them less than convincing. Do you have any trends of your own that you think support more modern societies?

Yes. Most of the graphs below come from 31 Charts That Will Restore Your Faith In Humanity.

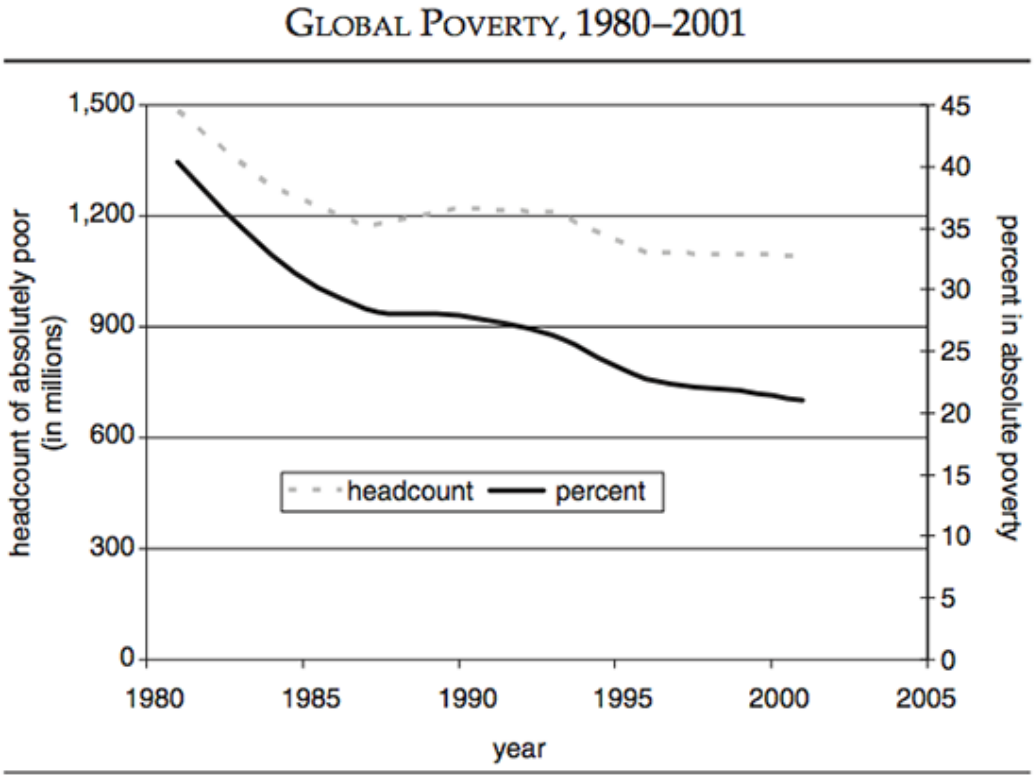


SOURCES: Angus Maddison, *Chinese Economic Performance in the Long Run* (Paris: OECD, 1998); Angus Maddison, *The World Economy: A Millennial Perspective* (Paris: OECD, 2001).

Hours worked per person

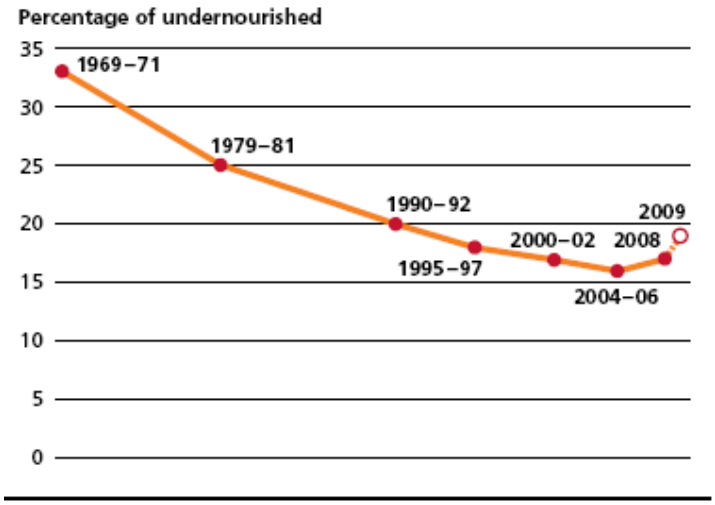


Global illiteracy



SOURCE: World Bank, *World Development Indicators*, http://devdata.worldbank.org/dataonline (accessed July 12, 2005).

Global poverty



World Hunger

I'm trying to keep things fair by deliberately excluding health care victories since these are at least partially due to technology, but these would include infant mortality dropping a hundredfold, the near elimination of smallpox, diphtheria, polio, tuberculosis, and typhoid from the developed world, the neutralization of AIDS.

Yet in reality, political and social trends played a role here too: for example, smallpox would not have been eliminated without the concerted effort of the WHO and other global health organizations.

1.9: Final thoughts on this section?

Of the seven categories Michael cites as especially supportive of the Reactionary thesis, zero are actually getting worse and several of them appear as best we can tell to be getting better. And I don't want to beat Michael up too much here, because these are the same sorts of things that other Reactionaries cite, and he got picked on only because he was the one to put them all in one place and claim he had evidence.

Reactionary claims that the modern world shows disappointing performance on indicators of social success turn out to be limited to one cherry-picked country or decade or else just plain made up. The very indicators Reactionaries cite turn out, on closer inspection, to provide strong evidence for things getting better.

Progressives, on the other hand, can point to some amazing victories over the last fifty years, including global poverty cut in half, world hunger cut in half, world illiteracy cut in half, war grinding almost to a halt, GDP quintuple-ing, violent crime collapsing, and self-reported happiness increasing in almost all countries.

1.9.1: Other than crime, few of these points have data before 1950, and the crime ones are highly speculative before that date. Don't you think that even if things have been getting better for the past few decades, they might have been getting worse over the past few millenia?

Yes. In a few cases this is obviously true. For example, Michael cites good data showing that traditional rural societies have lower suicide rates than our own. And obviously they have lower divorce rates. The same may be true with some of the other points here, though probably not as many as Reactionaries would like.

But I do think it's important to establish that things have been getting better over the past few decades. For one thing, it suggests a different course of action. If things are constantly declining, we should go into panic mode and try a radical restructuring of everything before it's too late. If things are getting better every day, we should hang tight and try to nudge forward trends that are already going on.

For another, it suggests a different interpretation. If things keep getting worse, we can attribute it to some process of social decay (since everyone seems to agree social decay is Getting Worse All The Time). If things are getting better now, we may perhaps separate societies into two groups, Traditional and Industrialized, admit that the transition from the first to the second caused a whole lot of problems, but be satisfied that industrialized society is gradually improving and fixing its defects.

So while I accept that traditional rural societies a thousand years ago were better on a number of social metrics, I don't think that's

particularly actionable. What's actionable is what's going on within industrial societies right now, and that seems to be improvements on all levels.

2: Are traditional monarchies better places to live?

2.1: Are traditional monarchs secure?

Much of the Reactionary argument for traditional monarchy hinges on monarchs being secure. In non-monarchies, leaders must optimize for maintaining their position against challengers. In democracies, this means winning elections by pandering to the people; in dictatorships, it means avoiding revolutions and coups by oppressing the people. In monarchies, elections don't happen and revolts are unthinkable. A monarch can ignore their own position and optimize for improving the country. See the entries on demotism and monarchy here for further Reactionary development of these arguments.

Such a formulation need not depend on the monarch's altruism: witness the parable of Fnargl. A truly self-interested monarch, *if sufficiently secure*, would funnel off a small portion of taxes to himself, but otherwise do everything possible to make his country rich and peaceful.

As Moldbug puts it:

> Hitler and Stalin are abortions of the democratic era – cases of what Jacob Talmon called totalitarian democracy. This is easily seen in their unprecedented efforts to control public opinion, through both propaganda and violence. Elizabeth's legitimacy was a function of her identity – it could be removed only by killing her. Her regime was certainly not the stablest government in history, and nor was it entirely free from propaganda, but she had no need to terrorize her subjects into supporting her.

But some of my smarter readers may notice that "your power can only be removed by killing you" does not actually make you more secure. It just makes security *a lot more important* than if insecurity meant you'd be voted out and forced to retire to your country villa.

Let's review how Elizabeth I came to the throne. Her grandfather, Henry VII, had won the 15th century Wars of the Roses, killing all other contenders and seizing the English throne. He survived several rebellions, including the Cornish Rebellion of 1497, and lived to pass the throne to Elizabeth's father Henry VIII, who passed the throne to his son Edward VI, who after surviving the Prayer Book Rebellion and Kett's Rebellion, named Elizabeth's cousin Lady Jane Grey as heir to the throne. Elizabeth's half-sister, Mary, raised an army, captured Lady Jane, and eventually executed her, seizing the throne for herself. An influential nobleman, Thomas Wyatt, raised another army trying to depose Mary and put Elizabeth on the throne. He was defeated and executed, and Elizabeth was thrown in the Tower of London as a traitor. Eventually Mary changed her mind and restored Elizabeth's place on the line of succession before dying, but Elizabeth's somethingth cousin, Mary Queen of Scots, also made a bid for the throne, got the support of the French, but was executed before she could do further damage.

Actual monarchies are less like the Reactionaries' idealized view in which revolt is unthinkable, and more like the Greek story of Damocles – in which a courtier remarks how nice it must be to be the king, and the king forces him to sit on the throne with a sword suspended above his head by a single thread. The king's lesson – that monarchs are well aware of how tenuous their survival is – is one Reactionaries would do well to learn.

This is true not just of England and Greece, but of monarchies the world over. China's monarchs claimed "the mandate of Heaven", but Wikipedia's List of Rebellions in China serves as instructional

(albeit voluminous) reading. Not for nothing does the *Romance of Three Kingdoms* begin by saying:

> An empire long united, must divide; an empire long divided, must unite. This has been so since antiquity.

.

Brewitt-Taylor's translation is even more succinct:

> Empires wax and wane; states cleave asunder and coalesce.

And of Roman Emperors, only about thirty of eighty-four died of even remotely natural causes, according to this List Of Roman Emperors In Order Of How Hardcore Their Deaths Were.

2.2: Are traditional monarchies more free?

A corollary of Reactionaries' "absolutely secure monarch" theory is that monarchies will be freer than democracies. Democrats and dictators need to control discourse to prevent bad news about them from getting out, and ban any institutions that might threaten the status quo. Since monarchs are absolutely secure, they can let people say and do whatever they want, knowing that their words and plans will come to naught. We revisit the Elizabeth quote above:

> Hitler and Stalin are abortions of the democratic era – cases of what Jacob Talmon called totalitarian democracy. This is easily seen in their unprecedented efforts to control public opinion, through both propaganda and violence. Elizabeth's legitimacy was a function of her identity – it could be removed only by killing her. Her regime was certainly not the stablest government in history, and nor was it entirely free from propaganda, but she had no need to terrorize her subjects into supporting her.

It is true that Elizabeth did not censor the newspapers, or bludgeon them into publishing only articles favorable to her. But that is less

because of her enlightened ways, and more because all newspapers were banned in England during her reign. English language news in the Elizabethan Era had to be published in (famously progressive and non-monarchical!) Amsterdam, whence it was smuggled into England.

Likewise, Elizabeth and the other monarchs in her line were never shy about killing anyone who spoke out against them. Henry VIII, Elizabeth's father, passed new treason laws which defined as high treason "to refer to the Sovereign offensively in public writing", "denying the Sovereign's official styles and titles", and "refusing to acknowledge the Sovereign as the Supreme Head of the Church of England". Elizabeth herself added to these offenses "to attempt to defend the jurisdiction of the Pope over the English Church…". Needless to say, the punishment for any of these was death, often by being drawn and quartered.

But at least she didn't have a secret police, right? Wrong. Your source here is Stephen Alford's book on, well, the Elizabethan secret police, although reason.com's review, The Elizabethan CIA: The Surveillance State In The 16th Century will serve as a passable summary.

2.2.1: How come we perceive traditional monarchies as less oppressive than for example Stalinist Russia?

Well, for one thing Stalin was in a category all of his own, going *far* beyond rational attempts to maintain his status into counterproductive paranoia. We shouldn't expect the average *communist police state* to be Stalinist in its intensity, and so we need not be surprised when traditional monarchies aren't.

But a more comprehensive answer might draw on a proverb of Oceania's in *1984*: "Animals and proles are free". Anyone too weak and irrelevant to be dangerous doesn't suffer the police state's attention.

Before about the 1600s, the average non-noble neither had nor could have any power. All wealth was locked up in land, owned by

nobles, and all military power was locked up in professionals like knights and men-at-arms, who could defeat an arbitrary number of untrained peasants without breaking a sweat.

After about the 1600s, wealth passed into the hands of capitalist merchants – ie non-nobles – and military power became concentrated in whoever could hold a gun – potentially untrained peasants. As a result, kings stopped worrying only about the nobility and started worrying about everyone else.

Or else they didn't. Remember, all of the longest and most traditional monarchies in history – the Bourbons, the Romanovs, the Qing – were deposed in popular revolts, usually with poor consequences for their personal health. However paranoid and oppressive they were, clearly it would have been in their self-interest to be more so. If monarchy were for some reason to be revived, no doubt its next standard-bearers would not make the same "mistake" as their hapless predecessors.

2.3: Are traditional monarchies less bloody?

Michael Anissimov writes:

> Bad kings are not nearly as bad as Demotist/Communist dictators. Bad kings are in a different universe from bad Demotist leaders. There is not even a vague comparison. In the traditional system, kings rely on the aristocracy and clergy for support, and have trouble doing anything without them. For a Demotist leader, there tends to be far fewer checks and balances. They can cause a million deaths in a place like Iraq with a snap of their fingers. Study up on the history of "death by government" to get a better perspective on what I mean. Kings and emperors very rarely, if ever, engage in mass murder against their own people.

2.3.1: Are demotist countries bloodier?

Look up demotist in a dictionary – Wiktionary will do – and you will find it means "one who is versed in ancient Egyptian demotic

writing". Mr. Anissimov's use is entirely idiosyncratic to Reactionaries, or, to put it bluntly, made up.

It is interesting that every time Reactionaries make this argument, they use this same made-up word. Here's Moldbug:

> Let's define demotism as rule in the name of the People. Any system of government in which the regime defines itself as representing or embodying the popular or general will can be described as "demotist." Demotism includes all systems of government which trace their heritage to the French or American Revolutions – if anything, it errs on the broad side.
>
> The Eastern bloc (which regularly described itself as "people's democracy") was certainly demotist. So was National Socialism – it is hard to see how Volk and Demos are anything but synonyms. Both Communism and Nazism were, in fact, obsessed with managing public opinion. Like all governments, their rule was certainly backed up by force, if more so in the case of Communism (the prewar Gestapo had less than 10,000 employees). But political formulae were of great importance to them. It's hard to argue that the Nazi and Bolshevik states were any less deified than any clerical divine-right monarchy.

Why use this made-up word so often?

Suppose I wanted to argue that mice were larger than grizzly bears. I note that both mice and elephants are "eargreyish", meaning grey animals with large ears. We note that eargreyish animals such as elephants are known to be extremely large. Therefore, eargreyish animals are larger than noneargreyish animals and mice are larger than grizzly bears.

As long as we can group two unlike things together using a made-up word that traps non-essential characteristics of each, we can prove any old thing.

None of Michael or Moldbug's interlocutors are, I presume, in favor of Stalinism or Nazism. They are, if anything, in favor of liberal democracies such as the United States or Great Britain. Michael and Moldbug cannot bring up examples of these countries killing millions of their own people, because such examples do not exist. So they simply group them in a made-up category with countries that have, and then tar the entire group by association. This is, of course, a riff on the good old Worst Argument In The World.

If there were any nonmotivated reason to group these countries together – if they were really taxonomically related – there would already be a non-made-up word describing this fact.

So the answer to the question – are demotist countries bloodier than monarchies? – is the same as the answer to the question "are eargreyish animals larger than grizzly bears". The answer is "Here's a nickel, kid; buy yourself a real category ."

2.3.2: Even if the "demotist" idea was invented for this debate, and even if it has little relevance to liberal democracies, isn't it at least a good basis for further study?

Remember Moldbug's definition: "Let's define demotism as rule in the name of the People. Any system of government in which the regime defines itself as representing or embodying the popular or general will can be described as demotist."

But "the leaders have to say they rule in the name of the people" is a pretty low bar. King Louis Philippe of France said he ruled in the name of the people:

> Louis-Philip wore the title of the King of the French…This title was in contrast to the King of France, which reflected a monarchy's power over the country, instead of a king's rule over its people. This title reflects that the king does not take his mandate from God but from the people themselves.

On the other hand, ever read *Les Miserables*? Yeah, that was him. Eventually the *actual* people hated him so much that they had a violent revolution and tried to kill him; the king managed to flee the capital in disguise and escape to England, where he died.

Why accept this stupid standard for the definition of "demotist"? Because a more reasonable one – like "elected by the people" or "liked by the people" or "not universally hated by the people and he has to have a giant army to prevent them from immediately killing him" would exclude for example Stalin, the figure Reactionaries are most desperate to paint as "demotist".

What about the regime which Reactionaries are the *second* most desperate to paint as "demotist"? For this one let's bring some class into this essay and quote Erik Maria Ritter von Kuehnelt-Leddihn:

> As an honest reactionary I naturally reject Nazism … fascism and all related ideologies which are, in sober fact, the reductio ad absurdum of so-called democracy and mob domination.

You heard it here first. The Nazis were *baaaaasically* the same as progressive liberal democrats.

To which all I can say is: you know who *else* opposed "so-called democracy and mob domination?"

> By rejecting the authority of the individual and replacing it by the numbers of some momentary mob, the parliamentary principle of majority rule sins against the basic aristocratic principle of Nature

– Adolf Hitler, *Mein Kampf*, p. 81

2.3.3: Even accepting all that, is Michael's last sentence even true?

Michael's argument ends by saying: "Kings and emperors very rarely, if ever, engage in mass murder against their own people."

I propose a contrary hypothesis – traditional absolutist regimes have always had worse records of massacre and genocide than progressives. However, technology improves efficiency in all things, including murder. And population has been growing almost monotonically for millennia. Therefore, it is unsurprising that more modern absolutist regimes – like Nazism and Stalinism – have higher death counts than older absolutist regimes – like traditional monarchies.

On the other hand, traditional monarchies have some *pretty impressive* records for killing their own people. Let us take a whirlwind tour of history:

The Albigensian Crusade, run by the French monarchy against its own subjects – with the support of the Catholic Church – may have killed up to a million people, which is pretty impressive considering that at the time there were only about twelve million Frenchmen. As a proportion of total population, this is about the same as the number of Germans who died during World War II, or Chinese who died during the Great Leap Forward.

The Harrying of the North was totally a real historical event and not something I stole from Game of Thrones. William the Conquerer, angry at the murder of a local earl, managed to kill about 100,000 northern Englishmen from 1069-1070, which was probably about 5% of the entire population.

Another 100,000 people died in the 16th century German Peasants' War, an event which so blended into the general mayhem of the time that you have never heard of it. Actually, the claim that Reactionary regimes have ever been peaceful would have trouble surviving a look merely at Wikipedia's *disambiguation* page for Peasants' War.

Third century BC emperor Qin Shi Huang was not only responsible for the Burning Of Books And Burying Of Scholars, but killed about one million out of his population of twenty million with

various purges and forced labor projects, one of which was the Great Wall of China.

[This section previously included a paragraph on Chinese warlord Zhang Xianzhong. Despite living in a 17th century monarchy, he held some pretty progressive values and his Reactionary credentials have been challenged. Rather than let his story distract from the more obviously Reactionary murderers above, I will concede the point]

But Michael goes even further. He says of democracies that "[with] a Demotist leader, there tends to be far fewer checks and balances. They can cause a million deaths in a place like Iraq with a snap of their fingers."

Ignoring for a moment the difference between snapping one's fingers and getting a bill to declare war passed through both houses of a hostile Congress (since Michael certainly does) we note that Michael has just authorized us to also compare monarchies and democracies in their ability to wreak havoc abroad.

On this particular historical tour, we will start with King Leopold of Belgium. Belgium itself was a constitutional monarchy run on a mostly democratic system, and in fact has always been a relatively pleasant and stable place. However, Belgium's colony, the Congo Free State, was under the direct rule of King Leopold. Not only was it responsible for the deaths of two to fifteen million Congolese – ie about as many Jews as were killed by Hitler – but the manner of those deaths was about as brutal and callous as can be imagined. Wikipedia writes:

> Leopold then amassed a huge personal fortune by exploiting the Congo. The first economic focus of the colony was ivory, but this did not yield the expected levels of revenue. When the global demand for rubber exploded, attention shifted to the labor-intensive collection of sap from rubber plants. Abandoning the promises of the Berlin Conference in the late 1890s, the Free State government restricted foreign access and

> extorted forced labor from the natives. Abuses, especially in the rubber industry, included the effective enslavement of the native population, beatings, widespread killing, and frequent mutilation when the production quotas were not met. Missionary John Harris of Baringa, for example, was so shocked by what he had come across that he wrote to Leopold's chief agent in the Congo saying: "I have just returned from a journey inland to the village of Insongo Mboyo. The abject misery and utter abandon is positively indescribable. I was so moved, Your Excellency, by the people's stories that I took the liberty of promising them that in future you will only kill them for crimes they commit."

This is an especially good example as it describes (we will see later) the ideal Reactionary state – one run by a single person identical to a corporation trying to make as much money as possible off a particular area and possessing overwhelming force.

The story does however have a happy ending – progressive elements within Belgium were so horrified that they forced the king to cede his claim – the colony was then governed by Belgium's democratically elected legislature, which did such a good job even Mencius Moldbug cannot resist the urge to praise it, and under whose rule Congo was a relatively liveable place up until a native uprising kicked out the Belgians and restored dictatorship.

Another good example of kings and emperors at war is Imperial Japan. This state – again run under principles no Reactionary could fault – accomplished the astounding feat of reducing the Nazis to the *second* biggest jerks on the Axis side during World War II. During the war, Imperial Japanese troops murdered between three million and ten million foreigners, mostly Chinese. Once again the brutality of their killings is impressive. According to Wikipedia on the Rape of Nanking:

> The International Military Tribunal for the Far East estimated that 20,000 women were raped, including infants and the elderly.[40] A large portion of these rapes were systematized in a process where soldiers would search door-to-door for young girls, with many women taken captive and gang raped.[41] The women were often killed immediately after being raped, often through explicit mutilation[42] or by stabbing a bayonet, long stick of bamboo, or other objects into the vagina. Young children were not exempt from these atrocities, and were cut open to allow Japanese soldiers to rape them

Meanwhile, Michael says that "Kings and emperors very rarely, if ever, engage in mass murder" but is *absolutely horrified* that America caused a million deaths in Iraq (more sober sources say 100,000, of which under 10,000 were civilians directly killed by US forces) while making the utmost effort to avoid unnecessary violence and launching war crimes proceedings against anyone caught employing it.

2.3.4: Conclusion for this section?

Reactionaries believe that monarchs are wise and benevolent rulers, and that it is only "demotists" who engage in genocide and mass murder.

But this argument is based on a con – "demotist" is an unnatural category they made up solely to win this debate. When we look at the governments their opponents actually support – liberal democracies – we find they have a much better history than monarchies.

Further, the Reactionaries fail *even on the terms of their own con*. Monarchs have a *fantastically* bloody history, and the regimes they want to paint as demotist really aren't.

2.4: Are traditional monarchs good leaders?

In his perhaps optimistically named "Ten Objections To Traditionalism And Monarchism, With Answers", Michael

Anissimov asks, with commendable bluntness: “What if the king is an idiot or psycho?” He answers:

> Then the prior king appoints a regent to take over the affairs of state on behalf of his successor. There is also a debate within the Reactionary community as to whether adoptive succession is preferable to hereditary succession, which avoids the issue of stupid or crazy children. Such extreme scenarios rarely ever happened during the age of Renaissance European monarchs. One of the greatest statesmen of all time, Klemens von Metternich, strongly influenced the mentally deficient monarch Ferdinand I of Austria during his reign, sat on the regency council, and ran most important affairs, presiding over a hundred years of relative peace in Europe.

We shall start with the theoretical objections before moving on to the empirical counterexamples.

Theoretical objection the first: what if the king doesn’t become an idiot or a psycho until after he is on the throne? The onset of schizophrenia can be as late as twenty-five; later in rare cases. Traumatic brain injury, certain infectious diseases, and normal human personality change can happen at any age. Smart psychopaths will have the presence of mind to avoid revealing their psychosis until they are safely enthroned.

Theoretical objection the second: what if the king seizes power some other way? A decent number of history’s monarchs got tired of waiting and killed their fathers. We would expect these to disproportionately include those who are crazy and evil, not to mention those who think their fathers would take away their power.

Theoretical objection the third: regency councils are historically about the least stable form of government imaginable. Unless everyone has truly commendable morality, either the king kills the regent and seizes power, the regent kills the king and starts a new dynasty, or some third party kills the regent and becomes the new

regent. Once again, reading *Romance of the Three Kingdoms* will prove instructional.

Theoretical objection the fourth: we are counting on the king's father to object if the king is an idiot or psycho. But a lot of idiotic psychotic kings' fathers were, in fact, idiots and psychos. The apple doesn't fall very far from the tree.

Onto the historical counterexamples. Historical counterexample the first: Gaius Julius Caesar Augustus Germanicus, "Caligula" to his friends. Absolutely beloved by the Roman populace. Unclear whether he killed his uncle Tiberius to gain the Empire, or just stood by cackling kind of maniacally as he died. Took power to general acclaim, ruled well for a couple of months, gradually started showing his dark side, and after a year or two reached the point where he ordered a large section of spectators at the colosseum to be thrown into the ring and torn apart by lions because the average amount of tearing-apart-by-lions at a Roman gladiatorial games *just wasn't enough for him*.

Historical counterexample the second: Ivan the Terrible. His father died of infection when Ivan was three years old. His mother was named as his regent – kind of a coincidence that the most qualified statesman in the realm would be his mother, but let's roll with it – but she died of poisoning when Ivan was eight. In this case I'm not sure who exactly is supposed to decide whether he's an idiot or psycho, and apparently neither were the Russians, because they crowned him Czar in 1547 . Ivan was okay until his wife died, at which point he became paranoid and started executing the nobility for unclear reasons, destroyed the economy, and burnt and pillaged the previously glorious city of Novgorod (part of his own kingdom!) with thousands of deaths. According to some sources:

> Ivan himself often spent nights dreaming of unique ways to torture and kill. Some victims were fried in giant frying pans and others were flayed alive. At times, he turned on [his death squads] themselves, and subjected their membership to torture

> and death. In a fit of rage, he murdered his own son; however the guilt of this act obsessed him and he never recovered.

Our story does not end there! Ivan died of a stroke, leaving the throne to his intellectually disabled son. Here at least the system worked – brilliant statesman Boris Godunov was installed as regent and ruled pretty well. He did, however, eventually seize the throne – likely because if he had not seized the throne everyone else would have killed him out of suspicion that he might seize the throne. He died, there was a huge succession squabble, and thus started the Time of Troubles, whose name is pretty self-explanatory.

Historical counterexample the third: Charles II Habsburg of Spain (not to be confused with various other Charles IIs). A strong contender for the hotly contested title of "most inbred monarch in history", Wikipedia describes him like so:

> Known as "the Bewitched" (Spanish: el Hechizado), he is noted for his extensive physical, intellectual, and emotional disabilities—along with his consequent ineffectual rule…
>
> Charles did not learn to speak until the age of four nor to walk until eight, and was treated as virtually an infant until he was ten years old. Fearing the frail child would be overtaxed, his caretakers did not force Charles to attend school. The indolence of the young Charles was indulged to such an extent that at times he was not expected to be clean. When his illegitimate half-brother Don Juan José of Austria, an illegitimate son of Philip IV, obtained power by exiling the queen mother from court, he covered his nose and insisted that the king at least brush his hair

As Charles's father died when Charles was 3, he was given a regent – his mother (*another* case in which the most qualified statesman in the land is the monarch's mother! What are the odds?!) But when his mother died, Charles took power in his own name and ruled for

four years. His only notable achievement during that time was presiding over the largest auto-da-fe in history. He died at age 39. Again quoting Wikipedia:

> The physician who practiced his autopsy stated that his body "did not contain a single drop of blood; his heart was the size of a peppercorn; his lungs corroded; his intestines rotten and gangrenous; he had a single testicle, black as coal, and his head was full of water." As the American historians Will and Ariel Durant put it, Charles II was "short, lame, epileptic, senile, and completely bald before 35, he was always on the verge of death, but repeatedly baffled Christendom by continuing to live."

Oh, and thanks to the vagaries of self-interested royal dynasties, his passing caused a gigantic succession struggle which drew in all the neighboring countries and caused hundreds of thousands of deaths.

Historical counterexample the fourth: Henry VIII. Really? Yes, really. While perhaps calling him an idiot or psycho goes too far, he certainly thought that marrying confirmed hottie Anne Boleyn and having a son with her was worth converting England to a newly-invented Protestant religion – a decision which killed tens of thousands, displaced some of the country's oldest and most important institutions, and set the stage for two hundred years of on-and-off warfare. Whether or not you like the Church of England (or, as it was almost named, Psychotic Bastard Religion) yourself, you have to admit this is a sort of poor reason to start a religious revolution.

King Henry wasn't an idiot or a psycho. He was just a selfish bastard. You can't expect his father to pick up on that. Even if you could, his father wasn't exactly Mahatma Gandhi himself. Worst of all, his personality may have changed following traumatic brain injury from a jousting accident – something that could not have been predicted before he took the throne.

This is exactly the sort of problem non-monarchies don't have to worry about. If Barack Obama said the entire country had to convert to Mormonism at gunpoint as part of a complicated plot for him to bone Natalie Portman, we'd just tell him no.

There's another important aspect here too. Reactionaries – ending up more culpable of a stereotype about economists than economists themselves, who are usually pretty good at avoiding it – talk as if a self-interested monarch would be a rational money-maximizer. But a monarch may have desires much more complicated than cash. They might, like Henry, want to marry a particular woman. They might have religious preferences. They might have moral preferences. They might be sadists. They might really like the color blue. In an ordinary citizen, those preferences are barely even interesting enough for small talk. In a monarch, they might mean everyone's forced to wear blue clothing all the time.

You think that's a joke, but in 1987 the dictator of Burma made all existing bank notes illegitimate so he could print new ones that were multiples of nine. Because, you see, he *liked* that number. As Wikipedia helpfully points out, "The many Burmese whose saved money in the old large denominations lost their life savings." For every perfectly rational economic agent out there, there's another guy who's *really* into nines.

2.5: Are traditional monarchies more politically stable?

Reactionaries often claim that traditional monarchies are stable and secure, compared to the chaos and constant danger of life in a democracy. Michael Anissimov quotes approvingly a passage by Stefan Zweig:

In his autobiography *The World of Yesterday* (1942), the writer Stefan Zweig described the Habsburg Empire in which he grew up as 'a world of security':

> Everything in our almost thousand-year-old Austrian monarchy seemed based on permanency, and the State itself was the chief guarantor of this stability . . . Our currency the Austrian crown, circulated in bright gold pieces, an assurance of its immutability. Everyone knew how much he possessed or what he was entitled to, what was permitted and what was forbidden . . . In this vast empire everything stood firmly and immovably in its appointed place, and at its head was the aged emperor; and were he to die, one knew (or believed) another would come to take his place, and nothing would change in the well-regulated order. No one thought of wars, of revolutions, or revolts. All that was radical, all violence, seemed impossible in an age of reason.

Michael's comment: "[This] does a good job capturing the flavor and stability of the Austrian monarchy…it's very interesting to read this in a world where America and Europe are characterized by political and economic instability and ethnic strife."

I am glad Mr. Zweig (Professor Zweig? Baron Zweig?) found his life in Austria to be very secure. But we can't just take him at his word.

Let's consider the most recent period of Habsburg Austrian history – 1800 to 1918 – the period that Zweig and the elders he talked to in his youth might have experienced.

Habsburg Holy Roman Austria was conquered by Napoleon in 1805, forced to dissolve as a political entity in 1806, replaced with the Kingdom of Austria, itself conquered again by Napoleon in 1809, refounded in 1815 as a repressive police state under the gratifyingly evil-sounding Klemens von Metternich, suffered 11 simultaneous revolutions and was almost destroyed in 1848, had its constitution thrown out and replaced with a totally different version in 1860, dissolved entirely into the fledgling Austro-Hungarian Empire in 1867, lost control of Italy and parts of Germany to revolts in the 1860s-1880s, started a World War in 1914, and was completely dissolved in 1918, by which period the reigning emperor's wife, brother, son, and nephew/heir had all been assassinated.

Meanwhile, in Progressive Britain during the same period, people were mostly sitting around drinking tea.

This is not a historical accident. As discussed above, monarchies have traditionally been rife with dynastic disputes, succession squabbles, pretenders to the throne, popular rebellions, noble rebellions, impulsive reorganizations of the machinery of state, and bloody foreign wars of conquest.

2.5.1: And democracies are more stable?

Yes, yes, oh God yes.

Imagine the US presidency as a dynasty, the Line of Washington. The Line of Washington has currently undergone forty-three dynastic successions *without a single violent dispute*. As far as I know, this is unprecedented among dynasties – unless it be the dynasty of Japanese Emperors, who managed the feat only after their power was made strictly ceremonial. The closest we've ever come to any kind of squabble over who should be President was Bush vs. Gore, which was decided within a month in a court case, which both sides accepted amicably.

To an observer from the medieval or Renaissance world of monarchies and empires, the stability of democracies would seem

utterly supernatural. Imagine telling Queen Elizabeth I – whom as we saw above suffered six rebellions just in her family's two generations of rule up to that point – that Britain has been three hundred years without a non-colonial-related civil war. She would think either that you were putting her on, or that God Himself had sent a host of angels to personally maintain order.

Democracies are vulnerable to one kind of conflict – the regional secession. This is responsible for the only (!) major rebellion in the United States' 250 year (!) history, and might be a good category to place Britain's various Irish troubles. But the long-time scourge of every single large nation up to about 1800, the power struggle? Totally gone. I don't think moderns are sufficiently able to appreciate how big a deal this is. It would be like learning that in the year 2075, no one even remembers that politicians used to sometimes lie or make false promises.

How do democracies manage this feat? It seems to involve three things:

First, there is a simple, unambiguous, and repeatable decision procedure for determining who the leader is – hold an election. This removes the possibility of competing claims of legitimacy.

Second, would-be rebels have an outlet for their dissatisfaction: organize a campaign and try to throw out the ruling party. This is both more likely to succeed and less likely to leave the country a smoking wasteland than the old-fashioned method of raising an army and trying to kill the king and everyone who supports him.

Third, it ensures that the leadership always has popular support, and so popular revolts would be superfluous.

If you remember nothing else about the superiority of democracies to other forms of government, remember the fact that in three years, we will have a change of leadership and almost no one is stocking up on canned goods to prepare for the inevitable civil war.

2.6: Are traditional monarchies more economically stable?

Once again, we come to Michael Anissimov's claims about Austria:

> Demotist systems, that is, systems ruled by the "People," such as Democracy and Communism, are predictably less financially stable than aristocratic systems. On average, they undergo more recessions and hold more debt. They are more susceptible to market crashes. They waste more resources. Each dollar goes further towards improving standard of living for the average person in an aristocratic system than in a Democratic one.
>
> The economic growth of the Austro-Hungarian Empire (1.76% per year) "compared very favorably to that of other European nations such as Britain (1%), France (1.06%), and Germany (1.51%)".

The growth of Austria-Hungary was higher than that of other European countries for the same reason the growth of sub-Saharan Africa right now is outpacing the growth of America or Europe – it was such a backwater that it had more room to grow.

Urbanization is a decent proxy for industrialization, and we consistently find that throughout the Kingdom of Austria and Austro-Hungarian Empire period, Austria had some of the lowest urbanization rates in Europe, just barely a third those of Britain, and well behind those of France, Spain, Italy, Germany, and Switzerland. In order to find a country as poorly developed as Austria-Hungary, we need to go to such economic powerhouses as Norway, Portugal and Bulgaria.

Nor was its economy especially stable. The Panic of 1873, probably the worst financial depression during the period being discussed and perhaps the worst modern economic crisis before the Great Depression, actually *started* in Austria-Hungary and only spread from there to the rest of the world. This is especially

astounding given Austria-Hungary's general economic irrelevance at the time.

2.6.1: What about Germany? Isn't the German Empire a good example of an industrially successful Reactionary country?

I consider the Reactionary credentials of the German Empire extremely open to doubt.

The German Empire was a utopian project created by people who wanted to sweep away the old patchwork system of landed nobility and local traditions that formed the Holy Roman Empire and turn it into a efficient modern state. The Progressive origins of both the Italian and German unification efforts shine through almost every word of a letter from Garibaldi to German unification pioneer Karl Blind:

> The progress of humanity seems to have come to a halt, and you with your superior intelligence will know why. The reason is that the world lacks a nation which possesses true leadership. Such leadership, of course, is required not to dominate other peoples, but to lead them along the path of duty, to lead them toward the brotherhood of nations where all the barriers erected by egoism will be destroyed. We need the kind of leadership which, in the true tradition of medieval chivalry, would devote itself to redressing wrongs, supporting the weak, sacrificing momentary gains and material advantage for the much finer and more satisfying achievement of relieving the suffering of our fellow men. We need a nation courageous enough to give us a lead in this direction. It would rally to its cause all those who are suffering wrong or who aspire to a better life, and all those who are now enduring foreign oppression.
>
> This role of world leadership, left vacant as things are today, might well be occupied by the German nation. You Germans, with your grave and philosophic character, might well be the ones who could win the confidence of others and guarantee

> the future stability of the international community. Let us hope, then, that you can use your energy to overcome your moth-eaten thirty tyrants of the various German states. Let us hope that in the center of Europe you can then make a unified nation out of your fifty millions. All the rest of us would eagerly and joyfully follow you.

The result of this idealistic vision – the destruction of the *ancien regime* in Germany – was a state much stronger than the traditional-but-weak Holy Roman Empire or anything that had existed in that part of the world before.

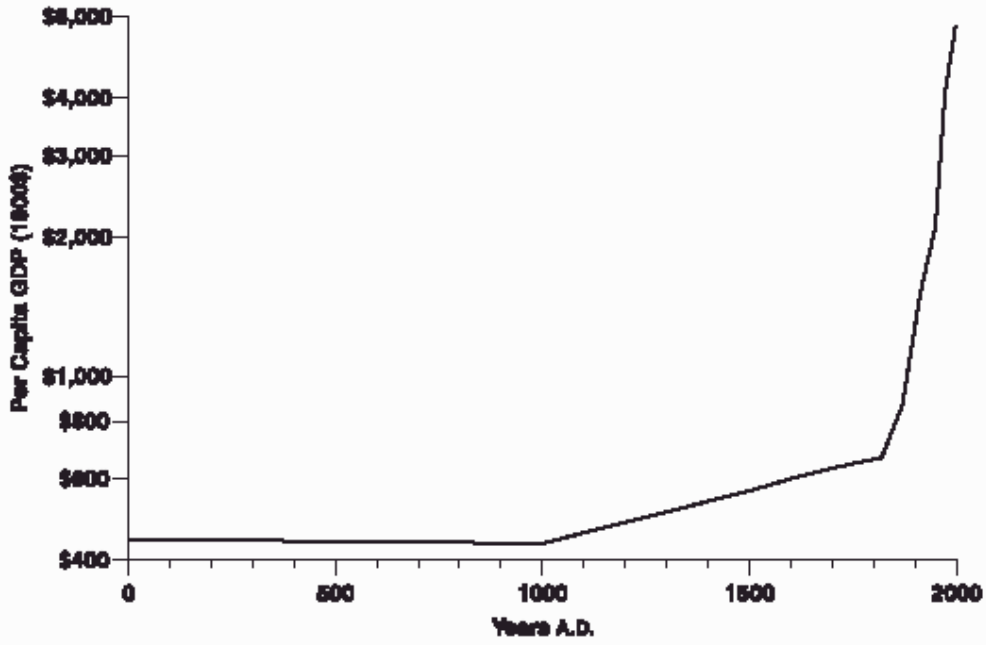
Sure, Otto von Bismarck was no hippie, but he was first and foremost a pragmatist, and his empire combined both conservative and progressive elements. It was based on a constitution, had universal male suffrage (only 5 years after the US got same!), elected a parliament, and allowed political parties. Granted, the democratic aspect was something of a facade to cover up an authoritarian core, but real Reactionaries would not permit such a facade, saying it will invariably end in full democracy (they are likely right).

The amazing growth of the German Empire was due to two things. First, the virtues of the German populace, which allow them to continue to dominate the European economy even today with an extremely progressive and democratic government. And second, the catch-up effect mentioned earlier. Germany had been languishing under traditional feudal and aristocratic rule for centuries. As soon as the German Empire wiped away that baggage and created a modern Progressive state, it allowed the economic genius of the Germans to shine through in the form of breakneck-speed economic growth.

2.6.2: Is Progressivism destroying the economy?

Another frequent claim. But remember how Michael said Progressivism went into high gear around the time of the French Revolution in 1789. Here's a graph of world GDP over time:

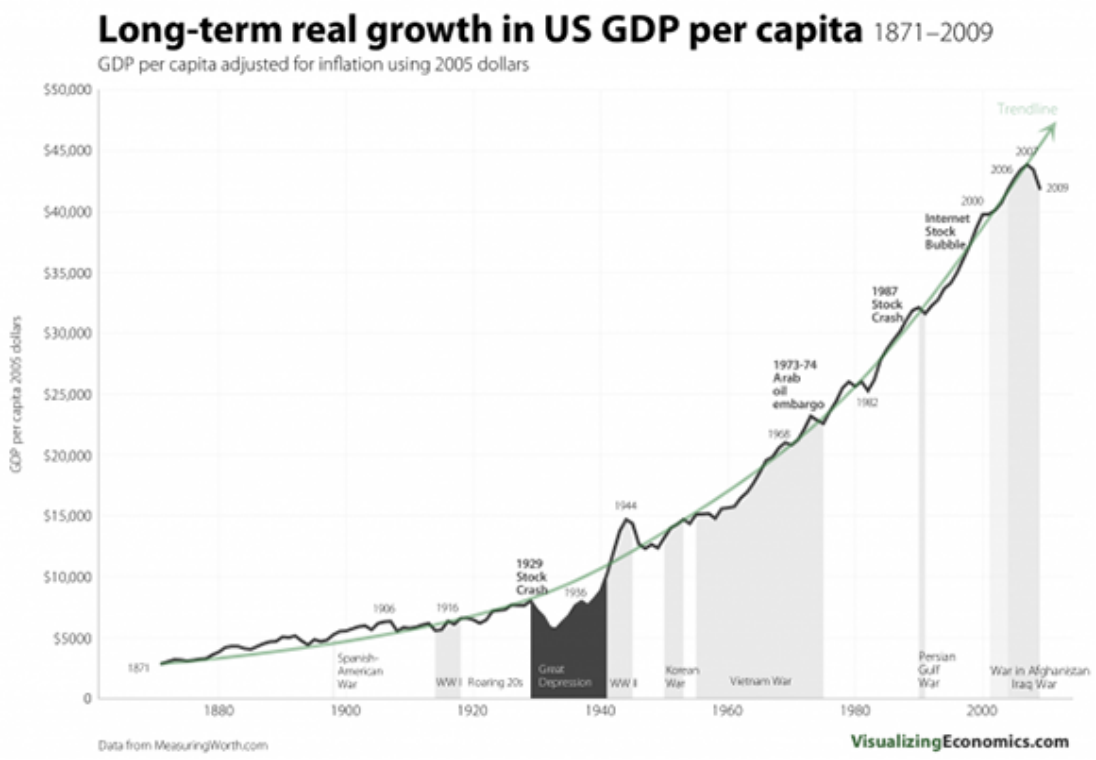
FIGURE 1–4 WORLD PER CAPITA GDP (INFLATION-ADJUSTED)



Source: Maddison, *The World Economy: A Millennial Perspective*, 264.

To put it lightly, I see no evidence of a decline starting around 1789?

Maybe the effect is just in the United States?



This image is actually even *more* astounding and important than the above, because it shows how growth keeps to a very specific trendline. On the graph above, the Reactionary might claim that technological advance was disguising the negative effects of Progressivism somehow. Here we see that *no* second variable that is not perfectly consistent has been interfering with the general economic growth effect.

I literally cannot conceive of a way that the data could be less consistent with the theory that Progressivism inhibits economic growth.

2.7: Are traditional monarchies just in general more successful and nicer places to live?

Great Britain and America have throughout their histories been the two most progressive nations on Earth. They've also been, over the past three hundred years or so, the two most successful. Other bright spots in the progressive/successful cluster include 1600s Netherlands, classical democratic Athens, republican Rome, and Cyrus' Persia. In fact, practically every one of the great nations of history was unusually progressive for its time period, perhaps with the exception of China – which is exceptionally complicated and hard to place on a Western political spectrum. Other possible exceptions might include Philip II's Spain, Louis XIV's France, and Genghis Khan's Mongolia – but the overall trend is still pretty clear.

Limiting our discussion to the present, our main obstacle to a comparison is a deficit of truly Reactionary countries. Reactionaries are never slow to bring up Singapore, a country with some unusually old-fashioned ideas and some unusually good outcomes. But as I have pointed out in a previous post, Singapore does little better than similar control countries, and the lion's share of its success is most likely due to it being a single city inhabited by hyper-capitalist Chinese and British people on a beautiful

natural harbor in the middle of the biggest chokepoint in the world's most important trade route.

Saudi Arabia also gets brought up as a modern Reactionary state. It certainly has the absolute monarchy, the reliance on religious tradition, the monoethnic makeup, the intolerance for feminist ideals, and the cultural censorship. How does it do? Well, it's nice and stable and relatively well-off. But a cynic (or just a person with an IQ > 10) might point out that a lot of this has to do with it controlling a fifth of the world's oil supply. It's pretty easy to have a good economy when the entire world is paying you bazillions of dollars to sit there and let them extract liquid from the ground. And it's pretty easy to be stable when you can bribe the population to do what you want with your bazillions of dollars in oil money – in fact, Saudi Arabia is probably that rarest of birds – a Reactionary welfare state.

(Actually, this point requires further remark. Reactionary states tend to be quite rich. In the case of Singapore, Reactionaries trumpet this as a success of Reactionary principles. In the case of Saudi Arabia, that sort of causation is somewhat less credible. I propose an alternative theory: Reactionary states can maintain themselves only by bribing the population not to revolt. These bribes may be literal, as in the case of the Saudi welfare state. Or they may be more figurative – "Look how rich my government has made you – you let me stay in power and I'll keep up the good work." China is the classic example of this particular formulation. This is important because contra Moldbug's inverted pendulum theory it suggests Reactionary regimes will be inherently unstable.)

But getting back to the issue at hand – given all these economic confounders, it's hard to compare Reactionary and progressive regimes in an even-handed way.

This is par for the course. Political science is notorious for its inability to perform controlled experiments, and no two countries will differ only in their system of government.

2.7.1: If we could perform a controlled experiment pitting reactionary versus progressive ideals, what would it look like?

Well, assuming you were God and had infinite power and resources, you could take a very homogeneous country and split it in half.

One side gets a hereditary absolute monarch, whose rule is law and who is succeeded by his sons and by his sons' sons. The population is inculcated with neo-Confucian values of respect for authority, respect for the family, and cultural solidarity, but these values are supplemented by a religious ideal honoring the monarch as a near-god and the country as a specially chosen holy land. American cultural influence is banned on penalty of death; all media must be produced in-country, and missionaries are shot on site. The country's policies are put in the hands of a group of technocratic nobles hand-picked by the king.

The other side gets flooded with American missionaries preaching weird sects of Protestantism, and at the point of American guns is transformed into a parliamentary democracy. Its economy – again at the behest of American soldiers, who seem to be sticking around a sufficient long time – becomes market capitalism. It institutes a hundred billion dollar project to protect the environment, passes the strictest gun control laws in the world, develops a thriving gay culture, and elects a woman as President.

Turns out this perfect controlled experiment actually happened. Let's see how it turned out!

Talk about your "Dark Enlightenment"!

From the Reactionary perspective, North Korea has done everything right. They've had three generations of absolute rulers. They've tried to base their social system on Confucianism. They've kept a strong military, resisted American influence, and totally excluded the feelings of the peasant class from any of their decisions.

Reactionaries, behold your god.

South Korea, on the other hand, ought to be a basketcase. It's replaced its native Confucian traditions with liberal Protestant sects, it's occupied by US troops, it's gone through various military coups to what the CIA calls a "fully functioning modern democracy", and it's so culturally decadent and degraded that it managed to produce *Gangnam Style*. Yet I don't think there's a single person reading this who doesn't know which one ze'd rather live in.

Yet according to the principles of Reaction (first quote Michael Anissimov, second Mencius Moldbug)

> Legally speaking, monarchies tend to have fewer laws, but enforce them more strictly, following Tacitus' dictum: "The more corrupt the state, the more numerous the laws." In general, monarchies put more power into the hands of local government. A key argument in favor of monarchy is that leaders tend to have a lower time preference, meaning they have a greater personal stake in the long-term well being of the country, compared to career politicians oriented towards four-year election cycles.
>
> A royal family is a family business. Not one king in European history can be found who ruined his own country to enrich himself, like an African dictator.

North Korea is a family business. And the Kim family has done very very well for itself. But it's not something I would like to see spread.

3: What is progress?

Reactionaries are not the first to notice – but may be the most obsessive in analyzing – a certain directionality to history. That is, rather than being a random walk across the space of possible values, at least the past three hundred years or so seem to have

shown a definite trend. Those who are in favor of this trend call it "progress". Those who oppose it call it things like "moral decay".

However, it is notoriously difficult to determine exactly what this trend is and what drives it. A theory to this effect is at the core of what separates Reactionaries from simple conservatives.

In the remainder of this section, I will replace the word "progress" – with its connotations of inevitability and desirability – with the preferred Reactionary term "progressivism" – that is, the political ideology which flows with the historical trend under discussion.

3.1: Might Progressivism be merely a secular strain of some Protestant religion?

Reactionaries seem to agree that Progressivism is a religion. Perhaps Calvinism. From Moldbug:

> I prefer "cryptocalvinism" [as a name for progressivism], meaning two things: that, like Calvin and as a direct result of his intellectual heritage, cryptocalvinists are building the Kingdom of God on Earth, a political system that seeks to eradicate every form of unrighteousness; and that they prefer not to acknowledge this characterization of their mission and heritage. Since I've changed the name, let me repeat the four ideals of cryptocalvinism: Equality (the universal brotherhood of man), Peace (the futility of violence), Social Justice (the fair distribution of goods), and Community (the leadership of benevolent public servants).

Or perhaps Quakerism. From Isegoria, quoting a different Moldbug theory:

> Modern progressivism is in fact a form of secular Quakerism, with its doctrine of the Inner Light only slightly modified.

Or how about Judaism? From Age of Treason:

> In a nutshell I object to [Moldbug]'s definition of Universalism, which is what he calls "the faith of our ruling caste". It's an important observation, but I think he gets it only half right. He associates Universalism only with Progressivism, which he blames entirely on Christianity. He does not address the Globalist tendencies of our ruling caste, and he pretty much gives Jews a pass…The close alignment of PC with Jewish interests? The Jewish support for Marxism and Bolshevism and hatred of Nazism perhaps?

Reactionaries seem much more certain that Progressivism is religious in origin than they are which religion exactly it originates from. And the differences between Calvinism and Quakerism are *not* subtle.

Given their total lack of consensus on a matter as basic as which religion, why is it so important to Reactionaries that progressivism be descended from a religious background? Moldbug explains:

> [Progressives] believe their ideals are universal, that they can be derived from science and logic, that no reasonable and well-intentioned person can dispute them, and that their practice if applied correctly will lead to an ideal society. I believe that they are arbitrary, that they are inherited from Protestant Christianity, that they serve primarily as a justification for the rule of the cryptocalvinist establishment, or Polygon, and that they are a major cause of corruption, tyranny, poverty and war.

So the reason Reactionaries want the Left to be religious is to disprove the contention that it is based on reason. This would presumably discredit the Left and restore preeminence to Reactionary ideas such as that people should be ruled by a king, live in strong heterosexual nuclear families, avoid sexual promiscuity, and derive their values from fixed traditions rather

than modern ideas of self-expression. You know, ideas with no religious background whatsoever.

3.1.1: Stop being snide and answer the question? Might Progressivism, far from deriving from some universal moral principles, actually be an arbitrary and parochial set of Calvinist customs and taboos?

The ideals commonly called progressive predate Calvin by several millennia. Consider the example of Rome. The early Romans not only overthrew their kings in a popular revolution and instituted a Republic, but experienced five plebian secessions (read: giant nationwide strikes aiming at greater rights for the poor). After the first, the Roman government created the position of tribune, a representative for the nation's poor with significant power in the government. After the third, the government passed a sort of bill of rights guaranteeing the poor protection against arbitrary acts of government. After the fifth, the government passed the Lex Hortensiana, which said that plebians could hold a referendum among themselves and *the results would be binding on the entire populace, rich and poor alike*. By the later Empire, even slaves were guaranteed certain rights, including the right to file complaints against their masters.

The Empire was remarkably multicultural, even at its very highest levels. Emperor Septimus Severus was half-Libyan and some historians think his appearance might have passed for black in modern America. Emperor Maximinus Thrax was a Goth, Emperor Carausius was Gallic, and Emperor Philip the Arab was…well, take a wild guess. Although Rome did have a state religion, they were extremely supportive of the rights of minorities to continue practicing their own religions, and eventually just tried to absorb everything into a giant syncretistic mishmash that makes today's "ecumenialism" seem half-hearted in comparison. Although their tolerance famously did not always extend as far as Christianity, when the Romans had to denounce it they claimed it was not a religion but merely a "superstition" – a distinction which itself

sounds suspiciously Progressive to modern ears. Indeed, the insistence of Christianity (and Judaism) on a single god, and their unwillingness to respect other religions as equally valid (in a very modern and relativistic way) was a large part of the Roman complaint against them.

The Romans pioneered the modern welfare state, famously memorialized by its detractors as *panem et circenses* – bread and circuses. Did you know welfare reform was a major concern of Julius Caesar? That ancient Rome probably had a higher percent of its population on the dole than modern New York? That the Romans *basically* worshipped a goddess of food stamps?

And no discussion of ancient Rome would be complete without mentioning their crazy sex lives. Wikipedia explains that "It was expected and socially acceptable for a freeborn Roman man to want sex with both female and male partners, as long as he took the penetrative role. The morality of the behavior depended on the social standing of the partner, not gender per se. Gender did not determine whether a sexual partner was acceptable, as long as a man's enjoyment did not encroach on another's man integrity." Gay weddings were not uncommon in ancient Rome, and were neither officially banned nor officially sanctioned. Juvenal and Martial both wrote satires condemning what they considered an epidemic of gay marriages during their era. And at least one Roman Emperor – Nero – married a man.

(well, married two men. One as groom and one as bride. And castrated one of them. And probably only married one of them because he was said to have an uncanny resemblance to Nero's mother. Whom Nero had previously had sex with, then murdered. I didn't say Nero was normal. Just unusually forward-thinking on the gay marriage issue.)

Moldbug listed the cryptocalvinist ie Progressive program as having four parts:

> Equality (the universal brotherhood of man), Peace (the futility of violence), Social Justice (the fair distribution of goods), and Community (the leadership of benevolent public servants)

Yet Equality has a clear antecedent in the plebian secessions of ancient Rome, peace in the Pax Romana, social justice in the Roman welfare system, and community in…well, it's so broadly defined here that it could be anything, but if we're going to make it the leadership of benevolent public servants, let's just throw in a reference to the philosopher-kings of Plato's *Republic* (yeah, fine, it's Greek. It still counts)

3.1.2: Yes, okay, the Romans tried to keep the peace and help the poor and stuff. That's a pretty weak definition of Progressivism. What really defines Progressivism is this messianic fervor that if we just do this *enough*, we can create a perfect utopia. *That* is what these ancient cultures were lacking. Even if you've never read *The Republic*, you can still get a sense of the utopian striving in the classical world from reading some of the stuff written during the reign of Emperor Augustus. Here's Dryden's translation of a passage from the *Aeneid*:

> An age is ripening in revolving fate
> When Troy shall overturn the Grecian state…
> Then dire debate and impious war shall cease,
> And the stern age be soften'd into peace:
> Then banish'd Faith shall once again return,
> And Vestal fires in hallow'd temples burn;
> And Remus with Quirinus shall sustain
> The righteous laws, and fraud and force restrain.
> Janus himself before his fane shall wait,
> And keep the dreadful issues of his gate,
> With bolts and iron bars: within remains
> Imprison'd Fury, bound in brazen chains;

> High on a trophy rais'd, of useless arms,
> He sits, and threats the world with vain alarms.

So please, tell me again how utopian desires for peace and social justice were invented wholesale by John Calvin in 1550.

3.2: Is the move toward Progressive social policy masterminded by "the Cathedral"?

Reactionaries have to walk a fine line. They can't just say "people consider liberal policies, decide they would be helpful, and form grassroots movements pushing for the policies they support", because that would make leftist policies sound like reasonable ideas pursued by decent people for normal human motives.

But they can't just say "There's a giant conspiracy where the heads of all the major Ivy League universities meet at midnight under the full moon", because that would sound ridiculous and tinfoilish.

So they invent this strange creature, the *distributed conspiracy*. It's not just people being convinced of something and then supporting it, it's them *conspiring to do so*. Not the sort of conspiring where they talk to one another about it or coordinate. *But still a conspiracy!* Michael Anissimov describes it like so:

> [The Cathedral is] the self-organizing consensus of Progressives and Progressive ideology represented by the universities, the media, and the civil service…the Cathedral has no central administrator, but represents a consensus acting as a coherent group that condemns other ideologies as evil […]
>
> Government and social policy is manufactured in universities, first and foremost at Harvard, followed by Princeton, then Yale, then the other Ivies, Berkeley, and Stanford. As far as politics is concerned, institutions outside of these are pretty much insignificant. Memetic propagation is one-way — it is formulated in the schools and pumped outwards. The universities are not significantly influenced by the outside.

> The civil servants that make government decisions are either borrowed from universities or almost totally influenced by them. The official mouthpiece of this ideological group is The New York Times, which is the most influential publication in the world outside of the Bible.

So now that we have this formulation of the problem, we can ask some more specific questions.

3.2.1: Are Harvard and the New York Times disproportionately linked to the Progressive ideas that now dominate society?

That depends partly on what "disproportionately" means, of course. But we can make some vague and qualitative observations.

The Roman and Persian Empires held some very Progressive ideals, all without the help of any universities or newspapers whatsoever. Parsimony suggests that whatever process pushed Rome to the left could be doing the same to the modern world.

But a better counterexample might be noting that even *modern* progressivism predates this institutions. The history of modern Progressivism – even as told by Reactionaries – goes from John Locke to the Glorious Revolution to the American Revolution to the French Revolution to the US Civil War on through John Stuart Mill to the New Deal and the United Nations and civil rights movements and on to the present. While Harvard (est. 1636) does predate all those events, I don't think even its most fervent critic would accord it any level of influence on world ideas until the 1850s at the earliest. And the Times was founded in 1851. It is hard to chart the precise progress of Progressivism, but I don't notice any sharp discontinuity at any point. Once again using parsimony, we might expect the forces that promoted Progressivism during the French Revolution and before to be the same forces promoting Progressivism afterwards. This takes any special role of Harvard or the New York Times entirely out of the pictures.

And modern progressivism doesn't seem linked to Harvard or the Times in *space* either. New York and Boston are pretty progressive – by American standards. But there's a whole world out there. Canada is further left than America; Britain is further left than Canada; France is further left than Britain; the Netherlands are further left than France; and Sweden is further left than the Netherlands. Russia and China are complicated, but they've certainly had their super-leftist periods. In fact, pretty much the entire developed world is further left than anywhere in the United States, New York and Boston not excepted. This does not seem an entirely recent development; for example, the Netherlands' liberalism has clear roots in the Dutch Golden Age of the 1600s.

It is true that sometimes a prophet is without honor in his own country. Yet for an American college and a newspaper read almost uniquely by Americans to have affected every other country in the Western world more effectively than they were able to affect the United States seems, well – unexpected.

3.2.2: Do Harvard and the New York Times invent Progressive dogma and then shove it down the throats of a hostile country?

Gay rights will be an interesting test here, because it's one of the issues on which society has shifted leftward most quickly and dramatically, and because it's relatively recent so its history should be easy to trace.

Modern gay rights movements trace their history to Germany, a country not known for having Harvard or the New York Times, or for that matter Puritans and Quakers. The German movement included such pioneering activists as Magnus Hirschfeld and Max Spohr, but Germany kind of dropped the ball on gay rights with the whole Nazi thing, and the emphasis shifted to elsewhere in Europe. In America, the movement finally gained steam in the 1960s with a picketing in Philadelphia and a community center in San Francisco, and finally the Stonewall Riots in New York.

I can't get any good information about Harvard's position, but the New York Times helpfully has an online archive of every article they have ever published. So what, exactly, was America's Newspaper Of Record doing while all this was going on? It was helpfully publishing articles like GROWTH OF OVERT HOMOSEXUALITY IN CITY PROVOKES WIDE CONCERN:

> The problem of homosexuality in New York became the focus yesterday of increased attention by the State Liquor Authority and the Police Department…The city's most sensitive open secret – the presence of what is probably the greatest homosexual population in the world and its increasing openness – has become the subject of growing concern of psychiatrists, religious leaders, and the police.
>
> Sexual inverts have colonized three areas of the city. The city's homosexual community acts as a lodestar, attracting others from all over the country. More than a thousand inverts are arrested here annually for public misdeeds. Yet the old idea, assiduously propagated by homosexuals, that homosexuality is an inborn, incurable disease, has been exploded by modern psychiatry, in the opinion of many experts. It can be both prevented and cured, these experts say.
>
> The overt homosexual – and those who are identifiable probably represent no more than half of the total – has become such an obtrusive part of the New York scene that the phenomenon needs public discussion, in the opinion of a number of legal and medical experts. Two conflict viewpoints converge today to overcome the silence and promote public discussion.
>
> The first is the organized homophile movement – a minority of militant homosexuals that is openly agitating for removal of legal, social, and cultural discriminations against sexual inverts. Fundamental to this aim is the concept that homosexuality is an incurable, congenital disorder (this is

> disputed by the bulk of scientific evidence) and that homosexuals should be treated by an increasingly tolerant society as just another minority. This view is challenged by a second group, the analytical psychiatrists, who advocate an end to what it calls a head-in-sand approach to homosexuality…

On and on and on it goes in this vein. And that's not even counting other such wonderful New York Times articles as WOMEN DEVIATES HELD INCREASING – PROBLEM OF HOMOSEXUALITY FOUND LARGELY IGNORED. These aren't editorials – this is the headlines, the supposedly fact-based objective reporting section. The editorials are worse – I particularly like the one warning that we need to fight increasing gay influence in the theater industry because gays cannot authentically write plays about love or relationships.

Now, to the Times' credit, it eventually changed its tune and is now mostly in favor of gay rights. That's fine for the Times but not so good for Reactionaries. The story here is very clearly of a gay rights movement that began as a grassroots push in favor of more tolerance. The New York Times opposed it, but *somehow* the movement managed to gather steam despite that crushing blow. Eventually its tenets became accepted by more and more people, and one of these late adapters was the New York Times, which now atones for its sin by defending gay rights against even *later* adapters.

This is not the pattern one would expect if all Progressive ideas were fueled solely by the New York Times' backing.

3.2.3: Do Harvard and the New York Times successfully impose their values on the rest of America and the world?

Let's examine exactly how opinions have changed on a host of important political issues. These are taken from the National Election Survey, Pew Research, and Gallup. I've tried to avoid cherry-picking – I took every issue I could find, starting from the

first year data was available. In cases where I could find two different polls, I kept the one with a longer data series:

Question	Original yea	% Then	% Now	Shift
Too much power in the hands of big companies	1987	77	77	0
Government should guarantee food and shelter	1987	62	62	0
I admire people who get rich working hard	1987	89	90	1
I am very patriotic	1987	89	88	1
We should fight for our country whether right or wrong	1987	54	53	1
We have gone too far pursuing equal rights	1987	42	41	1
Religion is important in my daily life	1982	37	36	1
Prayer is an important part of my daily life	1987	76	78	2
Labor unions have too much power	1987	59	61	2
I go to church at least once a week	1970	38	41	3
The federal government controls too much of our lives	1987	58	55	3
Businesses make too much profit	1987	65	62	3
Society should make sure everyone has = opportunity	1987	90	87	3
We should restrict immigration more than now	1987	76	73	3
Ban dangerous books from school libraries	1987	50	46	4
Dealing with federal government not worth the trouble	1987	58	54	4
What's good and evil always applies to all situations	1987	79	75	4
Get even with countries that take advantage of us	1987	44	49	5
Federal government should only run things local can't	1987	75	70	5
Government should help more needy people despite def	1987	53	48	5
I have a "pro-life" position on abortion	1970	56	51	5
Improve position of blacks with preferential treatment	1987	24	31	7
Poor are too dependent on gov assistance	1987	79	72	7
Need to be stricter laws to protect environment	1987	90	83	7
Gov should take care of those who can't care for selves	1987	71	63	8
Women should return to traditional roles	1987	30	19	11
People are responsible for getting jobs, not the govt	1970	49	61	12
I have old fashioned values about family and marriage	1987	87	71	16
We need stronger gun control	1995	60	44	16
I identify as a conservative	1970	21	40	19
Schools should have right to fire gay teachers	1987	51	28	23
Marijuana should be legal	1995	23	48	25
Gay people should be able to get married	1995	24	53	29
OK for blacks and whites to date	1987	48	83	35

Of thirty-four issues that made the cut, opinion shifted to the left on 19 and to the right on 13. There was an average shift of three points leftward per issue. Contrary to Reactionary claims that Americans do not appreciate the extent of the leftward shift affecting the country, in a recent survey based on a similar chart, most people, regardless of political affiliation, slightly overestimated the extent to which values had shifted leftward over the past generation.

Not only is the leftward shift less than people intuitively expect, it does not affect all issues equally. The left's real advantage is limited to issues involving women and minorities. Remove these, and opinion shifts to the left on 11 issues and to the right on 12. The average shift is one point rightward per issue.

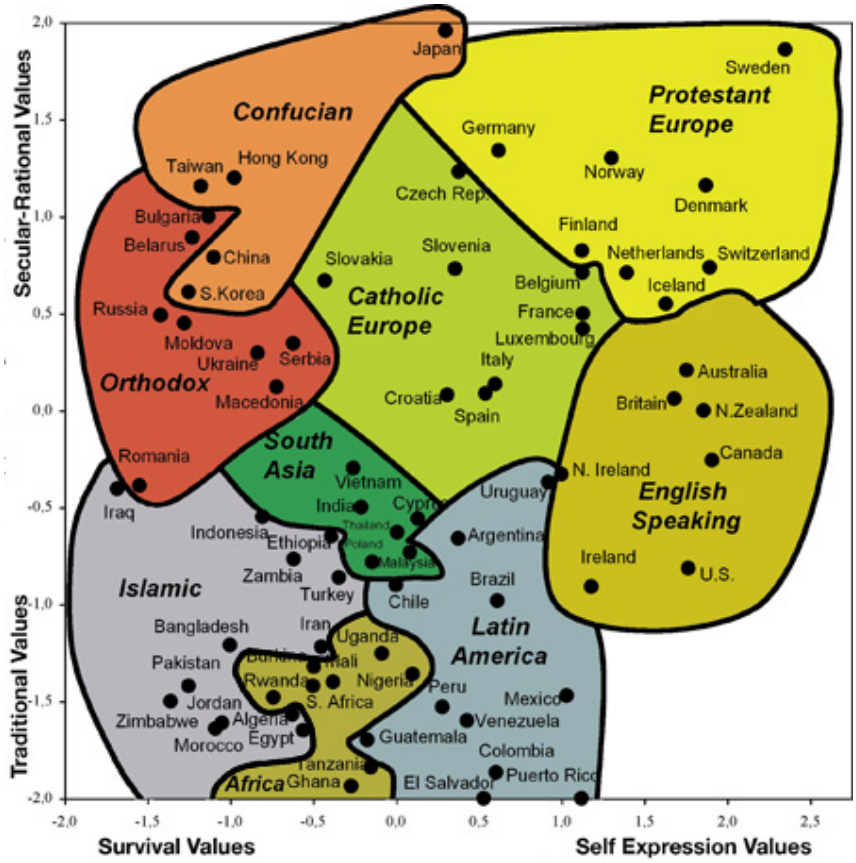
On the hottest, most politically relevant topics, society has moved leftward either very slowly or not at all. Over the past generation, it has moved to the right on gun control, the welfare state, capitalism, labor unions, and the environment. Although the particular time series on the chart does not reflect this, support for abortion has stabilized and may be dropping. This corresponds well with the DW-NOMINATE data that finds a general rightward trend in Congress over the same period. The nation seems to be shifting leftward socially but rightward politically – if that makes any sense.

If the Left had seized control of the government, or the media, or the institutions of the country, we would expect it to do a better job pushing its cherished policies like abortion rights, gun control, environmental protection, et cetera. Instead, beliefs on those issues have remained stable or shifted rightward, while issues like marijuana legalization – an issue more libertarian than progressive, and with minimal support from leftist institutions – succeed wildly. Whatever advantage the left has, it must be something skew to politics, something that institutionalized leftism, from the Democratic Party down to the Humanities Department at Harvard, can neither predict nor control.

3.3: Then where *does* progress come from?

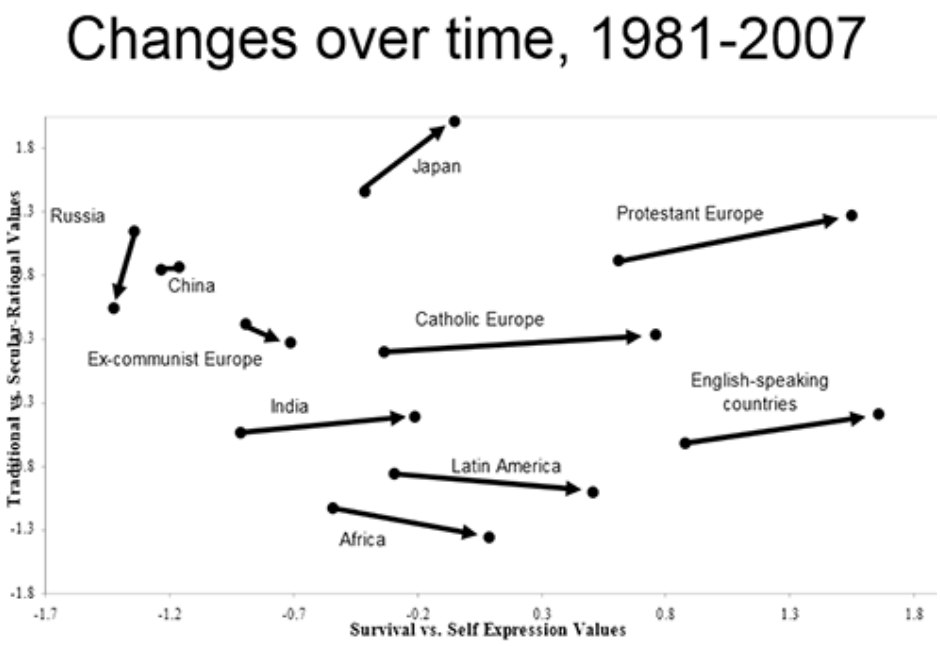
So the cultural shift of the past few centuries isn't toward some weird Christian sect. And it wasn't caused by Harvard or the New York Times. What was it and who did it?

The World Values Survey is the official academic attempt to understand this question. They've been polling in eighty countries around the world for thirty years trying to figure out who has what values and how they have been changing. Maybe you've seen the most famous summary of their results:

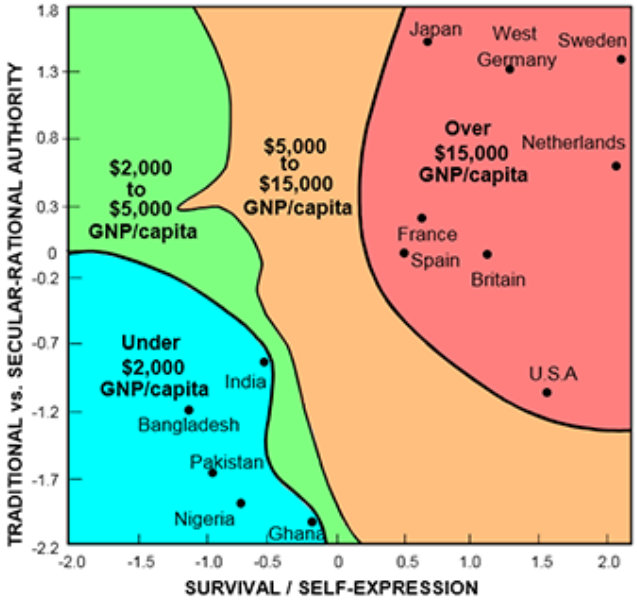


There is no end to the fun one can have with WVS data, and I highly recommend at least Wikipedia's Catalogue of Findings if not the original studies. But the most important part is that dimensionality analysis finds that answers to value questions cluster together onto two axes: survival vs. self-expression values, and traditional vs. secular-rational values.

Over time, societies tend to move from traditional and survival values to secular-rational and self-expression values. This is the more rigorous version of the "leftward shift" discussed above.

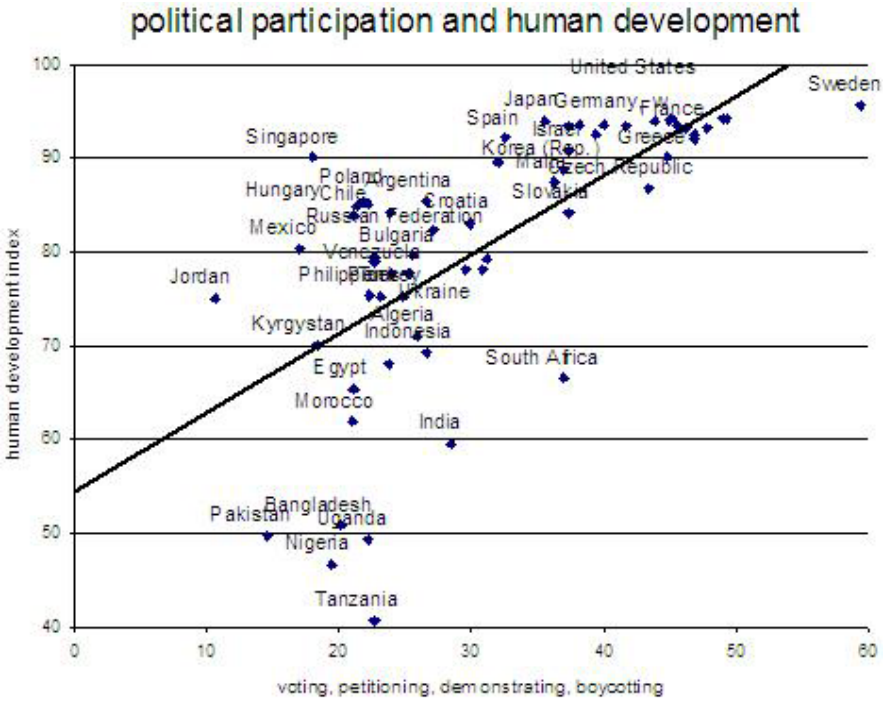


Both within a single time period and between time periods, traditional and survival values are generally associated with poverty, low industrialization, and insecurity. Secular-rational and self-expression values are generally associated with wealth, industrial or knowledge economies, and high security. The difference is not subtle:



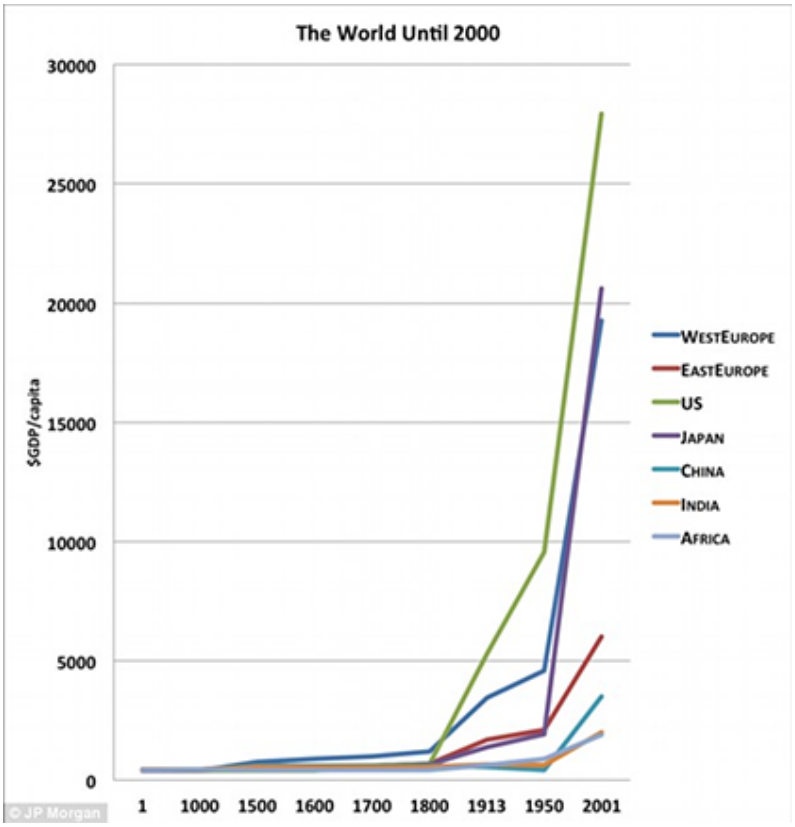
And if you want to know why countries are becoming more democratic and less monarchist, it's hard to get a more direct

answer than this graph (although its attempt at a linear fit was a bad idea):



All of this provides a simple and elegant explanation of the distribution of leftism, both in time and space. The most progressive countries today tend to be very wealthy, very peaceful, and comparatively urbanized. The least progressive countries tend to be poor, insecure, and comparatively rural.

Remember Michael Anissimov's description of the leftward shift above? That the world has been growing further to the left ever since the French Revolution? Take a look at the course of the world economy:



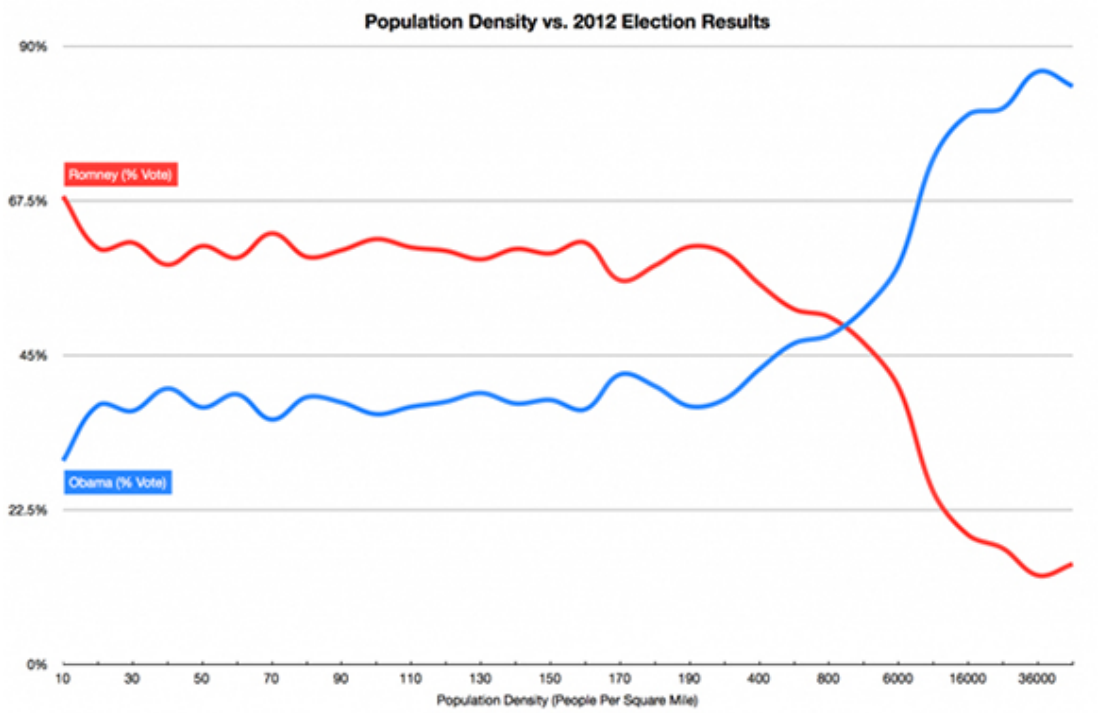
Riiiight about the time of the French Revolution – which also happens to be around the time of the Industrial Revolution – the world economy suddenly shifts into hyperdrive, starting in the USA and Western Europe, spreading to Japan after World War II, and not quite yet having reached Africa or Southeast Asia.

And, well, right about the time of the French Revolution Europe and the USA started shifting to the left, with Japan following after World War II, and Africa and Southeast Asia still lagging behind.

This progressivism/economics link is so obvious that anyone who thinks about it for a few minutes can reach the same conclusion. I wrote "A Thrive/Surive Theory Of The Political Spectrum long before I was familiar with the World Values Survey, but its conclusions match the survey's in pretty much every respect: rightist values are those most suited for hardscrabble existence where everyone must band together to survive a dangerous frontier; leftist values are those most suited for a secure postscarcity or near postscarcity existence with surplus resources available to devote to more abstract principles.

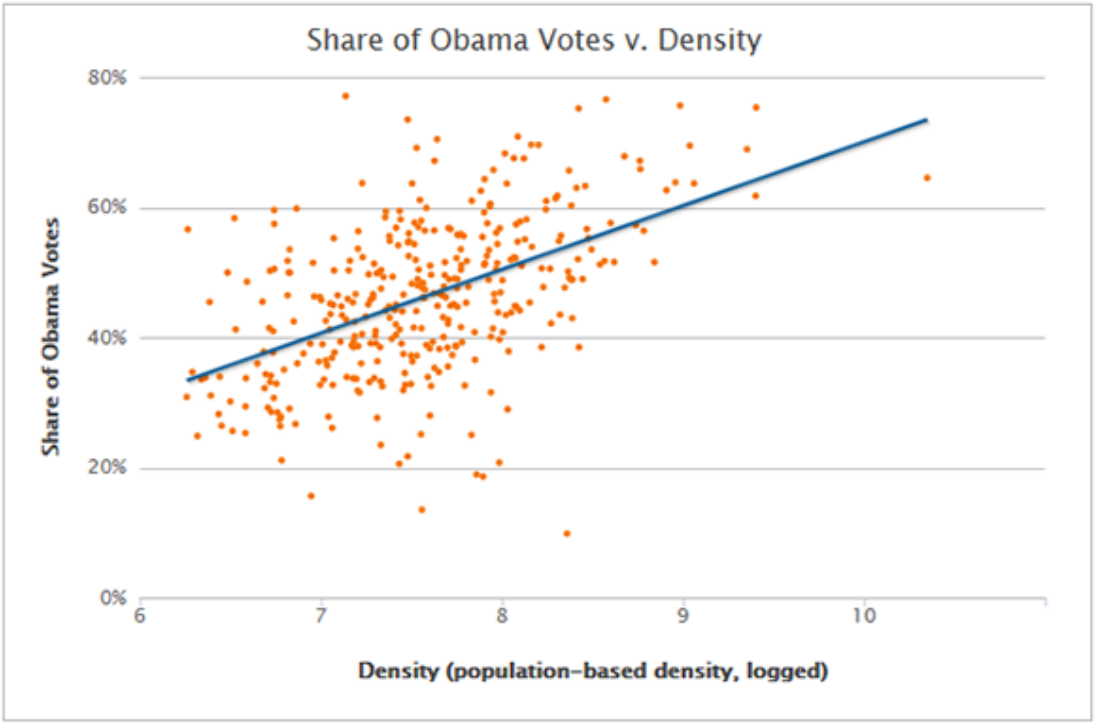
I'd like to examine one more aspect of this before I stop beating this dead horse, which is the rural/urban divide. The history of industrialization is in many ways the history of urbanization, and the distinction between insecure frontier life and secure postscarcity life mirrors the rural/urban divide. This predicts that more rural countries should be more traditional/survival and more urban countries more secular-rational/self-expression, which in fact we see. Of the countries furthest to the top-right on the WVS diagram, Sweden, Norway and Denmark all have about 85% urban populations. Go down to the three countries at the bottom left – Jordan, Morocco, and Zimbabwe – and despite Jordan's anomalously high level they're still averaging about 55%.

This is true even in the United States – the denser a county, city, or state, the more likely it is to lean Democratic, as we can see from [this terrible and confusing graph]:



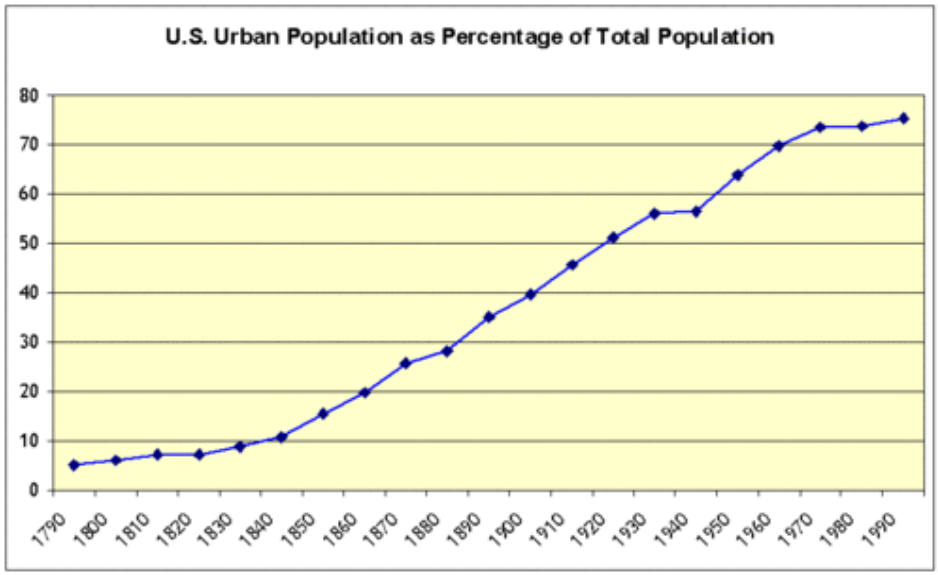
Rural counties – those with <200 people per square mile - lean red at about 65%. Once they pass that 200 person mark, they very quickly start leaning blue until the densest areas - true cities, approach 90% Democratic. Or as Dave Troy notes, "98% of the 50 most dense counties voted Obama. 98% of the 50 least dense

counties voted for Romney." This density effect applies even within cities. Here are America's largest cities graphed by density against percent Romney vote:



My sources point out that "graphs of the UK, Australia, and Canada look very similar during the same period, with left voting concentrated in urban and mining districts" and theyalso mention (just to fend off the inevitable reactionary critique) that "interestingly — and contrary to the much-stated view that Obama purchased the election with welfare, food stamps, and other entitlements, our analysis turned up no statistically significant association between Obama votes and the metro poverty rate and only a very small one for income inequality across metros."

Why am I making such a big deal of this? Well, here's America's percent urban versus percent rural population over the period of time when our values were shifting to the left:



So please. Tell me again how the leftward value shift over the past two hundred years was caused entirely by a sinister conspiracy of Ivy League college professors

3.3.1: Can you give a more detailed explanation of why increasing wealth, technology, and urbanization would lead to the values we call Progressive?

Here are five specific examples.

Multiculturalism is a forced adaptation to the culturally unprecedented situation of large groups of people from different cultures being forced to live and work together. This situation arises because of technology and urbanization. Technology, because more Somalis are going to immigrate to the US when that means booking a plane ticket over the phone than when it meant a six month journey over stormy seas. Urbanization, because it's much harder to immigrate into an agrarian society where every family knows each other and farmland is at a premium than into an urban society where you can apply for the same factory job as everyone else.

Modern gender roles are a forced adaptation to the existence of cheap and effective contraception, which decouples sex from pregnancy. Teen pregnancy is relegated to people unwilling or unable to use contraception, allowing other women to pursue the

same careers as men rather than dropping out of the workforce to become full-time mothers.

The welfare state is a forced adaptation to mobile and urban societies. In agrarian societies, most people owned their own means of production – their farms – and "unemployment" wasn't a salient concept. It was usually possible to get what you needed through the sweat of your brow, even if that meant chopping down trees to build a log cabin, and there was little sympathy for people who didn't bother. In urban societies, people need jobs in order to support themselves, and those who cannot get them starve in full pitiful view of everyone else.

Socialized health care is a very big part of the welfare state – probably the majority depending on how you parse the numbers. As recently as a century ago there really wasn't much in the way of health care technology for people to spend money on, and most people died quickly and simply without having to be kept alive in expensive hospitals for months. As health care gets beyond most people's ability to afford, and the average lifespan lengthens, there becomes more demand for government to step in and fill the gap.

Secularism is a more viable intellectual option once Science has discovered things like evolution and the Big Bang. Just as "there are no atheists in foxholes", people with a comfortable urban existence not dependent on the whims of the weather and the plague are less likely to worry about placating the Lord. Multiculturalism means that faiths no are no longer immune to challenge, as Christians and Muslims and Buddhists have to live next to each other and notice how totally unconvinced outsiders are of their ideas. And the movement from closely-knit communities to sprawling cities mean that the local church is no longer ties together your entire actual and possible social network so closely that it can exert pressure on you to conform.

And yes these are just-so stories, but the relationship between all these factors and wealth/urbanization are pretty much beyond

dispute – so if it's not true for these reasons it's true for reasons no doubt very much like them.

3.4: Do you believe in "Whig history"?

Whig history is an approach to historical study that emphasizes how the past has been groping towards the truths and institutions of the present. It is usually used derisively, in a sense of "Oh, so you think the era in which *you* were born just happens to be perfect, and everyone else from Aristotle to Galileo was just failing at being an American of 2013."

There is obviously a strong meaning of the term which cannot help but be false. The past did not share our values, it did not move linearly, and the present moment is neither perfect nor universally superior to other periods.

On the other hand, in a world where progress in areas as diverse as cars, computers, weapons and health care has been blindingly obvious, we shouldn't place too low a prior on the possibility that there has been progress in social institutions as well. Such progress could be motivated by the same factors that advance other areas.

First, a greater store of empirical results. As time goes on, we have more virtuous examples and terrible warnings. No one pushes for prohibition of alcohol anymore because we've seen how that turns out – and in thirty years, people may say the same about other drugs. Very few people push full hold-a-revolution Communism anymore, and for the same reason.

Second, better data. With the invention of statistics and information technology, we now have numbers on everything from income inequality to how different types of policing affect the crime rate. Members of the civil service, politicians, lobbyists, and even voters use these numbers to decide what policies to support. Neither the data nor its interpretation is always unbiased, but it's a heck of a lot better than the old method of doing whatever your prejudices tell you to do.

Third, social evolution. This is a complicated one, because all evolution is evolution to a niche, the niche is different in the modern world than in the medieval world, and so modern and medieval societies are optimizing for different things. But at the very least, we can say that modern institutions are better adapted to the modern niche than medieval institutions. Those governments that did not adapt were overthrown; those corporations that did not adapt went out of business; those institutions that did not adapt became unpopular and saw their influence shifted to other institutions. Those governments, corporations, and institutions that did adapt prospered and spun off copycats with small variations, and the evolutionary cycle repeated again.

To these three we could add things like greater education, better access to information, and more rational values (you can no longer get away with saying "Follow me because I'm the Messiah", and that's probably a good thing). So although it's not some a priori law of nature that the modern period must be the best period in history, we do have some reasons to expect things to be getting better rather than worse. As Part I pointed out, those expectations have mostly been realized.

3.5: Is America a communist country?

Reactionaries tend to push this line by finding the platform of the US Communist Party from some year well in the past, then pointing out that a lot of their goals were achieved, then noting that since America did what the communists wanted, we are a communist country.

Moldbug and others have claimed it, it even has its own Facebook page, but Free Northerner has done by far the most complete job analyzing it and finds that of demands in the 1928 Communist Party platform, 70% of all demands, and 78% of domestic demands, have been met as of 2013.

I don't want to belittle Free Northerner's work – he did a great job, he was much more rigorous than I'm about to be, and anyone who

writes a blog post on how awesome Turisas is is a friend of mine regardless of his political beliefs.

But although I can't get my computer to load the platform directly, I notice when I check his transcription that the Communist demands mysteriously lack points like "workers control the means of production" or "all property held in common", or even "not capitalism". They do, on the other hand, include policies like "abolition of censorship", "right to vote for everyone over 18", and "paid maternity leave during pregnancy".

Rather than conclude that America is a communist country, a better conclusion might be "the Communist Party of 1928 wasn't especially "communist", in the sense that we use that word today." That's no surprise. The meaning of words changes over time, and the Cold War made the more moderate elements of communism drop the "communist" label. Using a liberal definition of "communist" to claim that we satisfy the definition, then suggesting we should draw the conclusions and connotations we would from the strict definition of "communist" remains the worst argument in the world. Take out the Worst Argument In The World, and all the Communist Party platform experiment proves is that we support policies like "no child labor" and "free maternity leave" – ie things we already knew.

There's a second counterargument, though, which is more interesting. Free Northerner writes:

> I don't have time to analyze the Democratic and Republican platform demands of the same year at this time, but I would bet significant sums that less than 80% of their demands were met and upheld by our present time.

I'll take that bet!

I mistakenly got the Republican platform for 1920 (someone else can double-check 1928 specifically). The Republicans failed to

conveniently list their demands in bullet-point format, but from their long manifesto I managed to extract 37 different points:

1. Give farms right to cooperative associations
2. Protection against discrimination for farmers
3. End to unnecessary price fixing that reduces prices of farm products
4. Facilitate acquisition of farmland
5. Reduce frequency of strikes
6. Good voluntary mediation for industry
7. Convict labor products out of interstate commerce
8. Reorganize federal government
9. Simplify income tax
10. Federal Reserve free from political influence
11. Fair hours and good working conditions for railway workers
12. Private ownership of railroads
13. Immediate resumption of trade relations with all nations at peace
14. Restrict Asian immigrants
15. No one becomes citizen until they have taken a test to ensure they are American
16. American women do not lose citizenship by marrying an alien
17. Free speech, but no one can advocate violent overthrow of the government
18. Aliens cannot speak out against government
19. End lynching
20. Money for construction of highways
21. Save national forests and promote conservation
22. Reclaim lands
23. Increase pay of postal employees
24. Full women's suffrage in all states
25. Federal gov should aid states in vocational training
26. Physical education in schools

27. Centralize gov public health functions
28. End child labor
29. Equal pay for women
30. Limit hours of employment for women
31. Encourage homeownership for Americans
32. Make available information of housing and town planning
33. Americanize Hawaii
34. Home rule for Hawaii
35. Join international governing body such as League of Nations
36. No mandate for Armenia
37. Responsible government in Mexico

Not being too familiar with the 1920 political milieu, I don't really know what they mean by 2, 22, 32. Others seem so broad as to be hard to judge: 4, 6, 8, and 37. That leaves 29 points.

I think the Republicans have achieved 1, 3, 5, 7, 10, 11, 15, 16, 17, 19, 20, 21, 23, 24, 25, 26, 28, 29, 33, 34, 35, and 36 – some unambiguously, others if nothing else by the very sketchy criteria Free Northerner used to rule in commie achievements. They have definitely failed 9, 12, 13, 14, and 30. As for 18, 27, and 31, these seem ambiguous – let's count them half a point. That means they got 23.5/29 of the points they wanted – 81%. That's better than the Commies, who only got 70%.

(if we were really trying to do this right, we'd want to have the person who evaluated the success or failure of a party plank blinded to which party it came from. I'll leave someone else to try this).

So apparently the US is a Republican country even more than it's a Communist country. I bet if we looked over the Democratic platform for the same time frame, we'd find it was a Republican, Democratic, *and* Communist country. And if we check the Nazi Party platform, we find that some of the same points Free Northerner counts as Communist victories – abolition of child

labor, expansion of old age welfare – are also Nazi Party policies at the same time. So we are, in fact, a Democratic-Republican-Commie-Nazi country.

The alternative is that all parties liked to promise they would throw money at popular feel-good projects. Shorter working hours! Better welfare! Freedom of this! Freedom of that! As the country became richer it was able to support more feel-good policies, and so every party got much of what they wanted.

4: Could a country be ruled as a joint-stock corporation?

This is the plan of Mencius Moldbug, who gets points for being clever and creative rather than trying to rehash 13th century feudalism. I've heard different rumors as to whether he still supports it and whether it might or might not be a cover for supporting 13th century feudalism. Nevertheless the idea is interesting and deserves further investigation. However, it is missing some key details and suffers some probably irresolvable conceptual problems.

4.1: Would a joint-stock corporation prevent government decisions based on political tribalism and sacred values, in favor of government decisions based on maximizing economical value?

According to the theory, just as modern corporations like GE successfully remain dedicated to profitability, so America could be sold off in an IPO and restructured as a corporation dedicated to maximizing the value of US land.

But just calling something a corporation doesn't make it start worrying about profitability. Making its shareholders worry about profitability turns out to be surprisingly hard problem, even though these shareholders themselves would benefit from its profits.

We can imagine two different distributions of shares: either everyone gets one, or only a few aristocrats get one (the degenerate

third possibility, where only one person gets them, isn't really a "joint-stock company").

The first possibility might be suspected of being democracy: after all, every citizen equally has one share and therefore one vote. Moldbug argues it wouldn't be: shares are transferable, and citizens have an incentive to maximize the value of their share.

So chew on this: suppose that banning abortion would earn the American government $10 billion dollars a year (how? I don't know. Let's just say it does). This corresponds to about $30 for every American.

How many leftists do you think would vote to ban abortion for $30?

What if their $30 was entirely illiquid, only accessible by the one-time event of selling their single share of stock, and would probably be so lost in noise that they would never see tangible evidence of it?

Okay, what if they don't even *know* it will give them $30? No doubt Planned Parenthood will author a very scholarly report giving excellent reasons why an abortion ban will make stock shares plummet, and the Catholic Church will author an equally scholarly report giving excellent reasons why it will make everyone rich. Which side will people believe? Why, whichever side matches their natural prejudices, of course! As well ask a Democrat or a Republican whether Obamacare will increase or decrease the deficit.

The only thing that giving everyone a share of American stock would do to politics in the US is allow both the Left and the Right a chance to accuse one another of being secretly in it for the money, while both continue to do what they did before. Perhaps this wouldn't happen in a country created *de novo* out of thin air, but US politics are far too entrenched for giving people little stock certificates to help anything.

Anyway, it would take about ten minutes for poor people to sell their shares for easy cash. So this case would immediately degenerate to the second possibility – one where only a small "ruling class" owns all the stock certificates. I think a few Reactionaries have proposed this, and then they can be "nobles", and make up an "aristocracy", and…

Hold your horses. Suppose a new ruling class of ten thousand people possess all these certificates.

By definition all of these people will be multibillionaires – once you own one ten thousandth of America, you've got it made. And we observe something interesting with multibillionaires – Bill Gates, Warren Buffett, Larry Page. *They find other things much more interesting than money.* Bill Gates is working on curing malaria. Warren Buffett is trying to give all his money away to charity. Larry Page is working on fascinating but bizarre projects with minimal chance of success during his lifetime. Once you're a multibillionaire, you need more money less than you need to feel like you're making some kind of wonderful contribution to the world that will make coming generations revere you.

In other words, these shareholders won't care about the monetary value of their shares either. Take people like Ted Turner or the Koch brothers, give them a big chunk of the US government, and you expect them to focus on its *monetary value* just because you're calling it a stock?

4.2 Would corporate governance at least have lower discount rates?

Likely no.

Do corporations today have low discount rates? Consider the example of Lehman Brothers and other pre-crash investment banks. They happily accepted (and invented) subprime loans that would raise their profits today at the cost of likely financial disaster tomorrow.

More broadly, reflect upon how few companies pursue long-term revolutionary technology. Even though nearly everyone agrees that the future will be less based on fossil fuels, research and development of the likely replacements – from fusion power to solar power to electric cars – is either run by the government or grudgingly performed by corporations only after being promised huge government subsidies. When companies do develop exciting new technologies of their own accord – Google's Calico, SpaceX's rockets – they tend to be associated with some already-super-rich Silicon Valley mogul who has enough money to play around, rather than a sober corporation driven by the bottom line or investment opportunities.

A quick reflection on corporate incentives explains this pattern nicely. In the case of Lehman Brothers, traders got bonuses linked to year-on-year profitability, and because of coordination problems each had incentive to maximize his own bonus but no incentive to maximize the solvency of the company as a whole over time.

But why would a CEO or other corporate governor create such a structure? Well, although Reactionaries mock elected politicians for having a four-year time horizon, the average CEO stays only 6.8 years. That's less than a two-term president. And their *own* incentives are often *also* based on bonuses linked to short-term profitability.

In theory, the incentive to increase shareholder value ought to counteract short-term-ist tendencies. But it's an open question exactly how much of a time horizon is built into stock prices. The average investor holds the average stock for about seven months. Although the hope is that stock prices are set by the market discount rate, at an weighted average cost of capital of 10%, this ideal situation still means that anything happening thirty or more years from now determines only 4% of the stock price.

In the real world, it's even worse than this – CEOs have strong incentives to try to fool the market into short-term inflation of

stock prices at the cost of real future profitability. This is both common and successful. With many investors using formulae that extrapolate from past or present earnings to determine future earnings, it is unsurprising that the CEOs of companies like Lehman Brothers or Goldman Sachs were able to increase both their stock prices and their bonuses for many years until the inevitable letdown came – hopefully on someone else's watch.

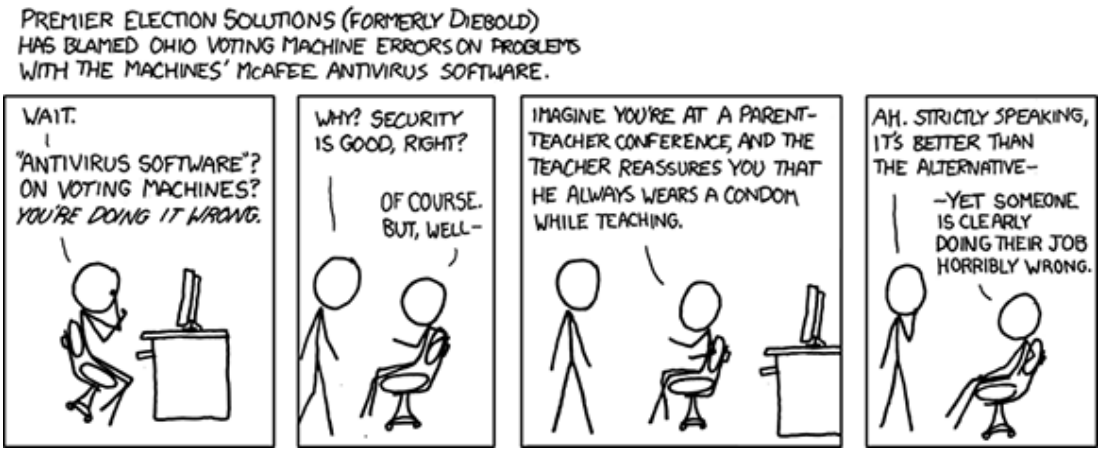
4.3: Could a joint-stock corporate state ensure complete security by mandating cryptographic locks on all its weapons?

This is one of Moldbug's proposals, and although I think it's been blown out of proportion and he's probably a little embarrassed by it now, it gets brought up enough to be worth addressing.

The idea is that shareholders of a corporate state possess cryptographic keys, and that these keys are necessary to fire the weapons in a country's arsenal. Therefore, any military coup can be stopped in its tracks.

The first question is exactly how these keys work. Suppose there are 100 shareholders. If all keys are necessary, then a single shareholder can paralyze the military. If 51 of the 100 keys are necessary – well, I don't know if cryptography can implement such a scheme securely, but let's suppose that it does.

One can raise some peripheral problems with this scheme. Having all your country's guns connected to the Internet might not be such a good idea…



…and it would be sort of unfortunate if your entire military could be brought down by a clever hacker or Scott Aaronson building a quantum computer in his basement. Further, the guns would have to be either default-on or default-off. If they were default-on, then military conspirators could disable the communications network (or just the radios on their weapons) and have free rein. If they were default-off, then a foreign military could disable the communications network and take over the country because none of the military's weapons would work.

More important, this only protects against a small subset of rebellions. If every unit has a separate code, it may be able to give loyalist military units the advantage over treasonous units in the case of intramilitary feuding. But it can't can't stop a popular revolution – the type where rebels become guerillas and gradually defeat the military in combat. It happened in China, it's happening right now in Syria, and it could happen again regardless of any cryptographic locks on weapons.

4.4: Would shareholder value maximization be a good proxy for making a country a nice place to live?

Suppose that all the above problems are solved, and we have installed a genuinely self-interested CEO with a long time horizon. Will the new policy of increasing shareholder value really make the country a nicer place as well?

In many ways the equivalence holds. If, as Moldbug suggests, a corporate state's profits came from land value taxes, and so profits came from increasing land values, then things like decreasing crime, pollution, and poverty would be in the corporate state's best interests. So would allowing its residents enough freedom to make moving to its land attractive.

But the ways it *doesn't* hold are really horrible.

Businesses have an incentive to please their paying customers. As Mitt Romney informs us, a large proportion of Americans don't pay taxes. In fact, they consume government resources in the form

of welfare, while providing no economic value in return. In some cases, these citizens are "fixer-uppers", people who with enough investment could become productive. In other cases – the indigent elderly, the physically and mentally handicapped, or just people with no useful skills – keeping them around would be a poor financial decision. When regular companies find they have people who aren't producing value, they "downsize" them. It's unclear what exactly would be involved in "downsizing" unproductive American citizens, but I'm betting it wouldn't win any Nobel Peace Prizes.

In a post called The Dire Problem And The Virtual Option, Moldbug discusses some of these problems with his system. He admits that this is a major issues (the titular "dire problem"). With his trademark honesty:

> As the King begins the transition from democracy, however, he sees at once that many Californians – certainly millions – are financial liabilities. These are unproductive citizens. Their place on the balance sheet is on the right. To put it crudely, a ten-cent bullet in the nape of each neck would send California's market capitalization soaring – often by a cool million per neck. And we are just getting started. The ex-subject can then be dissected for his organs. Do you know what organs are worth? This is profit!

But his proposed solutions are bizarre and in many cases incomprehensible:

> The simplest, broadest, and most essential prevention against this degenerate result is the observation that the royal government is a government of law, and a government of law does not commit mass murder. For instance, no such government could take office without promising to preserve and defend its new subjects, certainly precluding any such genocide.

A government of law is different from a “law-abiding citizen” or “law-abiding business” in that governments, in addition to occasionally following the law, also get to *make* the law. If the government had some strong incentive to shoot citizens, it could pass a law allowing it to shoot citizens. It is no more than dozens of other governments have done throughout history. Such a law need not even ruffle the feathers of its more productive “assets”: it could come up with some very clear criteria for whom to shoot and then stick to those criteria scrupulously.

No government could take office without promising to preserve and defend its new subjects *in a democracy*. Or, to be broader, no government could take office under such conditions as long as it was responsible to its populace and depended on their support. The entire premise of Moldbug’s utopia is a government whose rule is by force and does not depend on the consent of the governed.

If Moldbug’s King needed to gain the consent of the governed before taking power, they wouldn’t stop at making him sign a promise not to shoot anyone. They would make him sign a promise to rule for the good of the people rather than in order to maximize shareholder value. Heck, the last time we tried something like this, the people made the government sign the Bill of Rights.

Here Moldbug wants to have his cake and eat it too. His government will be unconstrained and effective because it doesn’t rule by consent of the people. But when we start examining how horrible an “unconstrained effective government” really would be, he promises that need for the consent of the people would rein it in.

Positing a government that can ignore the age-old constraint of popular consent is far-fetched enough. Positing one where the constraint *only* arises in those situations where it would be optimal for it to arise, but not otherwise, is just dreaming.

> But do we really know it? The explanation that Royal California will not harvest the poor for their organs, because it will have promised not to harvest the poor for their organs,

> and its most valuable asset is its reputation, while certainly accurate, is too narrow for me. Having established this legalistic defense, let us reinforce it with further realities. More broadly, Royal California will in all cases treat her subjects as human beings. The maintenance of equity, as well as law, is crucial to her reputation. Thus, the Genickschuss is out, with or without the organ harvesting.

Our second layer of protection is that the king will preserve human rights and maintain equity among persons. I wonder if the person writing this has ever read Mencius Moldbug. He has some pretty interesting arguments against human rights and the equity of persons, and I'd be interested in hearing a debate between the two of them.

> Carlylean to its core, the ideology of Royal California is that the King is God's proxy on earth; whatever God would have him do, that is justice; the King, having done his best to divine God's will, shall see it done. Or else he is no king, but a piece of cardboard, a "Canadian lumber-log." Clearly, God is not in favor of harvesting the poor for their organs. You're probably thinking of Huitzilopochtli. So this is another safeguard.

So our third layer of protection – and I am not making this up – is "the will of God". Don't you feel safer already? Politicians would never do bad things, even when it is in their own self interest, because God wouldn't want them to. I think that's pretty much all the protection citizens might need from their government, don't you? Let's write a letter to the libertarians and tell them they can all go home now, *God* has this one covered.

But I should not be too harsh on Moldbug. He goes on to admit we probably do need a fourth layer of protection, beyond the three he has mentioned. And he even steel-mans the case against him, noting that in a higher-technology world, more and more people will become unproductive until, instead of being a tiny proportion

of citizens, it may become the majority or (in the post-Singularity case) everyone who has to worry about this. He gives a few possible solutions:

> First, the King has no compunction whatsoever in creating economic distortions that produce employment for low-skilled humans. A good example of such a distortion in the modern world are laws prohibiting self-service gas stations, as in New Jersey or Oregon. These distortions have gotten a bad name among today's thinkers, because makework is typically the symptom of some corrupt political combination. As the King's will, it will have a different flavor.
>
> As both a good Carlylean and a good Misesian, the King condemns economism – the theory that any economic indicator can measure human happiness. His goal is a fulfilled and dignified society, not maximum production of widgets. Is it better that teenagers get work experience during the summer, or that gas costs five cents a gallon less? The question is not a function of any mathematical formula. It is a question of judgment and taste. All that free-market economics will tell you is that, if you prohibit self service, there will be more jobs for gas-station attendants, and gas will cost more. It cannot tell you whether this is a good thing or a bad thing.
>
> There may be no jobs for men with an IQ of 80 in Royal California – at least, not in a Royal California whose roads are paved by asphalt rollers. But suppose its roads are paved in brick? A man with an IQ of 80 can lay brick, do it well, and obtain dignity from the task. Nothing whatsoever prevents the King from distorting markets to create demand for the supply he has.

Okay, so the corporate CEO in a government based solely on maximizing shareholder value will decide to trash his own economy in order to provide jobs for the jobless, because that's just

how much corporate CEOs respect human dignity. This is just like corporate CEOs today, who never fire anyone to increase profitability because maintaining jobs is more important. Sure, let's roll with that.

Since we have abandoned the free market here, we no longer have the free market's safeguards on job tolerability. Depending on how many make-work jobs the King creates, we will have either an oversupply, an undersupply, or a just-right-supply of unskilled laborers to fill them, which in turn will determine workers' wages and living conditions. Will the King maintain them at a living wage in good conditions, or at conditions more like the immigrant farm laborers of today? If the latter, I *suppose* that's better than killing off the unproductives, but it's still pretty dystopian. If the former, then that's quite nice of the King, but I can't help noting that by instituting useless make-work government-provided jobs for everyone at guaranteed salaries, he has kind of just re-invented Communism, which seems to be the sort of thing I would have expected Reactionaries to try to avoid.

I would compare this idea to the idea of a Basic Income Guarantee. Both cost the economy the same amount of money. Yet in Moldbug's plan, the poor spend their entire day digging ditches and filling them in again. In a basic income guarantee, the poor spend their days doing whatever they want – producing art, playing games, or working to make themselves more productive. Moldbug may wax rhapsodic about the dignity of work, and he is not entirely wrong, but the sort of work that has dignity is not the sort of work where you dig ditches and fill them in again to earn a government-set paycheck. I wonder if you asked the employed gas station attendant and the unemployed bohemian to rate the level of dignity they feel they have, would this support Moldbug's thesis?

But never fear, Moldbug has yet another plan:

> Or not. The low-browed man of 70 (and remember – for every 130, there is a 70) may still require special supervision.

> Besides a job, he needs a patron. Productivity he has, but direction and discipline he still requires. His patron may be a charity, or a profitable corporation, or even – gasp – an individual.
>
> In the last case, of course, we have reinvented slavery. Gasp! Since the bond of natural familial kindness is not present in the case of an unrelated ward, the King keeps a close watch on this relationship to protect human dignity. Nonetheless, his wards are farmed out – it is always better to be a private ward than the ward of the State. Bureaucratic slavery is slavery at its worst. Adult foster care, as perhaps we will call it, is a far more human and dignified relationship.

So, we will force people to work for other people against their consent, but it will all be okay and humane, because the government will be keeping "a close watch on this relationship"? Darnit, I liked it better when we were being protected by "the will of God".

If Moldbug agrees that bureaucratic slavery is "slavery at its worst", what exactly does he mean when he says the King will "keep a close watch" on these "adult foster care" institutions. Will the King personally go out to each of them and evaluate? That seems like a lot of work in a state of 40 million people. Or will he appoint some government officials to do so, to inspect each institution and make sure it is up to code? If so, how is this different from "bureaucratic slavery"? Is it because the bureaucrats and slaveowners aren't *literally* the same people?

Look, Moldbug. I know you don't think you're reinventing Communism, *but you are*.

Luckily he has one more trick up his sleeve:

> If a human being cannot support himself in a civilized manner in the King's economy, which has been carefully tweaked to match labor demand to labor supply, the King does not

> provide a “safety net” in the 20th-century style, in which he may lounge, sag, bob and fester forever. No – then, it is time for the Virtual Option.
>
> If you accept the Virtual Option – always a voluntary decision, even if you have no other viable options – California will house, feed and care for you indefinitely. It will also provide you with a rich, fulfilling life offering every opportunity to obtain dignity, respect and even social status. However, this life will be a virtual life. In your real life, your freedom will be extremely restricted: to the point of imprisonment. You may even be sealed in a pod.
>
> The result is that the ward (a) disappears from society, and (b) retains or (hopefully) increases his level of dignity and fulfillment. He remains a financial liability, because it is still necessary to prepare his meals and maintain his pod. But other residents of California no longer feel menaced by his presence. For he is no longer present among them.

This doesn’t sound so bad to me, although I’m probably a huge outlier on this and if you actually tried it on people you’d have a civil war on your hands.

But first of all, it’s impossible with current levels of technology, always a bad sign.

Second of all, it’s something that would be equally viable in a democracy and a monarchy. Compare these pods to television. Right now, we pay welfare money to the poor, and, in some cases, they use that money to watch television all day. When they complain, it generally is not due to a lack of television but to a lack of money. If we had virtual reality pods, no doubt the situation would look little different, and conservatives and Reactionaries would be the ones complaining that we pay the poor money to sit in virtual reality pods all day instead of getting a real job.

Third of all, it would probably cost more than any other option. Putting a man in prison – feeding him, boarding him, and putting some guards on the doors to make sure he doesn't escape costs about $50,000 a year – more than sending that same man to any college in the country. The bulk of the expenses are health care and security – two problems that would be equally dire in these pods. In fact, solving the medical problems associated with prolonged immobility in a virtual environment might be further beyond our current technology than the virtual environment itself.

If the true reason behind the Virtual Option is keeping the poor out of everyday society – even though many of its residents would be old people, disabled people, and the like – why not just offer those people $40,000 a year to live in some nice community out in the country made up solely of other non-working poor? It would be cheaper, more humane, and after a few years with a stable income and a normal life the people involved might end up being unexpectedly productive.

This is, of course, a question one could ask of our own society as well as of Moldbug's hypothetical. So let's stick to criticizing Reactionaries, which is more fun and less depressing.

4.5: Would exit rights turn countries into business-like entities that had to compete with one another for citizens?

Exit rights are a great idea and of course having them is better than not having them. But I have yet to hear Reactionaries who cite them as a panacea explain in detail what exit rights we need beyond those we have already.

The United States allows its citizens to leave the country by buying a relatively cheap passport and go anywhere that will take them in, with the exception of a few arch-enemies like Cuba – and those exceptions are laughably easy to evade. It allows them to hold dual citizenship with various foreign powers. It even allows them to renounce their American citizenship entirely and become sole citizens of any foreign power that will accept them.

Few Americans take advantage of this opportunity in any but the most limited ways. When they do move abroad, it's usually for business or family reasons, rather than a rational decision to move to a different country with policies more to their liking. There are constant threats by dissatisfied Americans to move to Canada, and one in a thousand even carry through with them, but the general situation seems to be that America has a very large neighbor that speaks the same language, and has an equally developed economy, and has policies that many Americans prefer to their own country's, and isn't too hard to move to, and almost no one takes advantage of this opportunity. Nor do I see many people, even among the rich, moving to Singapore or Dubai.

Heck, the US has fifty states. Moving from one to another is as easy as getting in a car, driving there, and renting a room, and although the federal government limits exactly how different their policies can be you better believe that there are very important differences in areas like taxes, business climate, education, crime, gun control, and many more. Yet aside from the fascinating but small-scale Free State Project there's little politically-motivated interstate movement, nor do states seem to have been motivated to converge on their policies or be less ideologically driven.

What if we held an exit rights party, and nobody came?

Even aside from the international problems of gaining citizenship, dealing with a language barrier, and adapting to a new culture, people are just rooted – property, friends, family, jobs. The end result is that the only people who can leave their countries behind are very poor refugees with nothing to lose, and very rich jet-setters. The former aren't very attractive customers, and the latter have all their money in tax shelters anyway.

So although the idea of being able to choose your country like a savvy consumer appeals to me, just saying "exit rights!" isn't going to make it happen, and I haven't heard any more elaborate plans.

5: Are modern ideas about race and gender wrongheaded and dangerous?

The past century has seen a huge opening up of racial and sexual norms, as a closed-minded traditional society willing to dismiss everything against their personal morals as disgusting or evil started first discussing and later embracing alternative ideas.

This was followed by a subsequent closing back up of those norms, as society decided it was definitely right this time, and this time *for real* anyone who brought up any alternative possibilities was definitely disgusting and evil.

Reactionaries deserve kudos for lampshading these taboos and pointing out various modern hypocrisies in a frank and honest way. But to invert an old saying, I will defend to the death their right to say it, but disagree with what they say.

5.1: Are modern women sluts?

This is a surprisingly important question in Reactionary thought. Just to prove I'm not strawmanning:

> So you might say, Bryce, if you want an objective and useful definition of the word slut, you would have to conclude that most Western women are sluts. That's not good. And I say "Exactly."

– Anarcho-Papist

> Obviously democracy is not working, is failing catastrophically. The productive are outvoted by the gimmedats, in large part non asian minorities and white sluts.

– blog.jim.com

> Why would you take a slutty girl seriously? Once she accepted slut into her life, keep her out of yours. It is rare for a slut to truly reform so I would not even take the chance. Once

> a slut, always a slut. Do you really want your kids coming out the same place 10 other men have gone into? "But doesn't that pretty much rule out about 85% of women or so?" Well, unfortunately it does. I wish there was a better answer but there is not. Do not settle for sluts, if they have such little respect for themselves imagine how little respect they will have for you. Manning up does not mean settling for a hopeless graying slut."

– Occidental Traditionalist

> We live in strange times. Recently several religious conservative bloggers have suggested that the word "slut" is a slur against all women, and that it is a type of profanity. My best guess is they feel that sluts know that what they are doing is wrong, so even using the word in general is cruel to their already convicted hearts.

– Dalrock

Telling women that sleeping around is bad just because it's "slutty" is argument through mere connotation of words. Then again, accusing these people of "sexism" or "misogyny" would be the same. So let's bury the insults and try to figure out what's going on.

Are people becoming sluttier? Several studies have addressed this question (though, uh, not in those exact words). In America, we have only a few scattered studies recording a shift from an average of two lifetime sexual partners for women and six for men in 1970 to about four partners for women and six for men in 2006. But we change methodologies midstream and have to confuse means with medians to get those numbers. France is the only country to do the study properly, perhaps unsurprising given their legendary love of all things amorous. Their numbers seem similar to ours but more precise, so let's use the French results:

> Number of partners reported in the lifetime remained stable between all three surveys for men of all ages (11.8 in 1970, 11.0 in 1992, and 11.6 in 2006). For women, mean lifetime number of partners increased from 1.8 in 1970 to 3.3 in 1992 and to 4.4 in 2006.)

One of the first things we notice about these data is that they cannot possibly be true. Men cannot be having more (heterosexual) sex than women, nor can the two statistics trend in different directions. The least mathematically impossible explanation is that between 1970 and 2006, women have become less likely to lie about all the sex they're having.

Does that contradict common sense, which tells us everyone is really slutty nowadays but was perfectly chaste in the past? Maybe, but common sense seems to be not entirely correct. Common sense would tell us that modern young people are having much more sex than youth fifteen years ago, but according to the study "no increase was observed between 1992 and 2006 in women under thirty; for men under thirty a decrease in the mean was seen in the most recent period – 10.4 in 1992 and 7.7 in 2006, $p < 0.00001$" (the growth of the Muslim population in France from 7% to 10% during that time period seems insufficient to account for the changes) **5.1.1: If a woman is a slut, does that mean her future marriage is doomed to failure?**

Before you answer, consider a common failure mode. Some rule catches on for some very useful reason. Like "don't have sex with your cousin, you'll have kids with two heads." Biological or memetic evolution selects for people who follow the rule, and eventually the rule becomes an unquestionable taboo.

But historically no one understood Mendelian genetics. The rule didn't make sense, but it had to be followed. And so people came up with rationalizations. Some of them were simple rationalizations for simple folk: "don't have sex with your cousin, God hates it." Or "Don't have sex with your cousin, it's disgusting." More

sophisticated people demanded more sophisticated rationalizations: eventually you get “Don’t have sex with your cousin, it could go wrong and damage the structure of trust necessary for an extended family”, or “Don’t have sex with your cousin, it is contrary to this here complicated conception of natural law”.

Then suppose the original reason for the rule is taken away. Someone wants to have protected sex with their cousin, understanding that they cannot ethically have children. Or someone invents a gene therapy that allows people to have sex with their cousins without additional risk of birth defects.

Doesn’t matter. Everyone will have had so much fun making up rationalizations that they will object to the new harmless act almost as much as to the old dangerous act. “God still hates it!” “It’s still disgusting!” “It still damages the family structure of trust!” “It’s still contrary to the natural law!”

But it would be very strange if, the original reason for the belief having been neutralized, by coincidence the belief happens to be right anyway. Imagine that an explorer comes back from a distant jungle with a tale of a humongous monster. Everyone catches monster fever and begins speculating on how the monster may have gotten there. Then the explorer admits his tale is a hoax. Objecting “But there could still be a monster there!” is fruitless. If the original reason anyone held the belief is invalid, it’s unlikely that by coincidence the belief just happens to be correct.

Let’s get back to sluttiness. (I am following the lead of my interlocutors in concentrating on female sluttiness only here, since it seems to be the only type anyone cares about. Yes, you’re very clever for pointing out that men can be promiscuous as well. Why don’t you follow it up with the phrase “double standard” or a reference to “playing the field”?)

We know two *very* good reasons why sluttiness has been stigmatized in nearly all societies. First, slutty women were more likely to get sexually transmitted diseases. Second, slutty women

were likely to end up with children outside of wedlock. Back when men were the sole providers and didn't have much providing to spare, that would have been just about a death sentence.

These are two *huge* issues. These two issues alone are more than sufficient to explain the taboo on sluttiness establishing itself on every continent and in every major religion. These are more than sufficient to explain why some people think sluts are disgusting, why they're low status, why we have a cultural taboo on sluttiness.

But of course, most sluts today have these two issues figured out. Contraception prevents the out of wedlock births. Protection and antibiotics prevent the STDs. So the old reasons no longer hold.

It would be quite the coincidence if a taboo that formed for one reason *just happened* to be vitally important for society for totally different reasons.

I admit the Reactionaries have their justifications for why sluttiness is bad. They say sluttiness before marriage can lead to sluttiness after marriage, and thither to infidelity, divorce and broken families. Or the slut's previous experiences might have given her higher expectations, leading to divorce and broken families again. And…

…no, that's actually all the justifications I can find. There are people who think they have other justifications, but they can never explain them in so many words. Read this article. No, really, read that article. Gods! Have you ever seen so many mere assertions and Arguments From My Opponent Believes Something in one place?

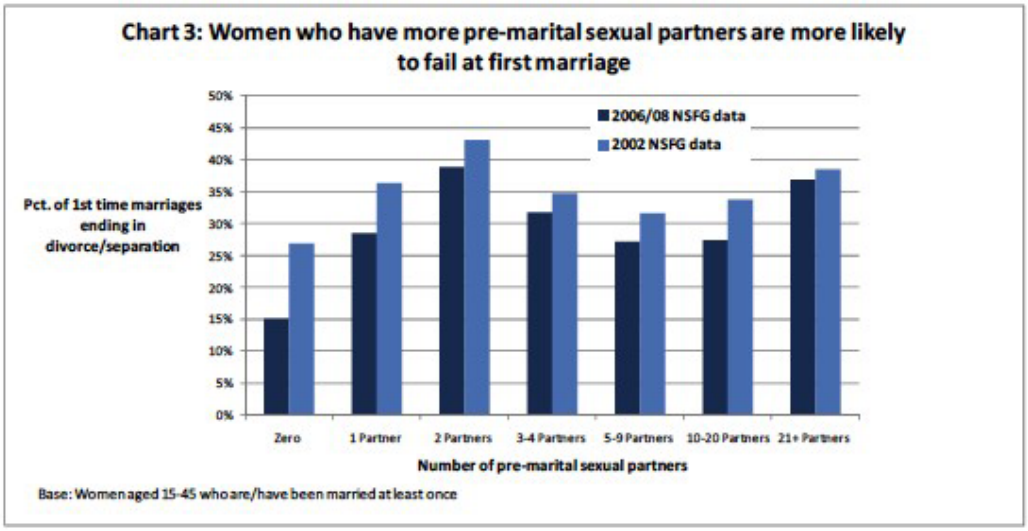
So okay. They have two just-so stories. I can come up with just-so stories too! Like – if a woman sleeps with a lot of people before marriage, she'll be better able to estimate how compatible she is with any given partner. Or – if a woman can sleep with men before marriage, she won't be compelled by horniness to marry the first loser she meets just so she can have sex with someone. Or – if a woman has a couple of relationships before she marries, she'll have practice with relationships and won't screw the important one up.

This is fun! How about – if a woman sleeps with people before settling down, she won't feel curiosity that makes her stray afterwards?

The reason these sorts of just-so stories about sluttiness keep popping up is the disappearance of the good historical arguments against the practice, leaving behind only a feeling of disgust in search of a justification.

One might argue – isn't the proof in the pudding? Divorce rates have been going up lately, infidelity rates have been going up; correlation isn't always causation but isn't it at least suggestive?

In this case, no. We can even check. From Social Pathology:



Women with zero or one premarital sexual partners have more stable marriages than women with two or more partners. Okay. Who gets married a virgin these days? Super-religious people. They're not going to divorce. And from the source, I gather that most of these stably married one partner women are women who had premarital sex with their future husband. Super-religious people who slipped up. Their poor self-control earns them a 15% lower likelihood of stable marriage: harsh, but fair.

The people with two or more partners are the ones who we know are "experimenting" – having sex with at least one person other than their future husband. Among this group, likelihood of unstable marriage goes *down* with more partners up until you reach the 20

partner or so level – at which point you're probably capturing prostitutes, cluster B personality disorders, and other people outside the mainstream.

The data provide some evidence that an absolute commitment to purity – no sex before marriage, or sex only with your husband-to-be – predicts marital stability. But beyond that – in the two to twenty partner range in which recent social change has been occurring – there's no correlation between increasing sluttiness and decreasing marital stability.

5.1.2: Woman only put out for macho but antisocial men. Our society encourages that tendency and shames "beta males" who are nice and prosocial but cannot get women. This incentivizes men to become jerks, and men follow those incentives in droves. Don't we need to do something about women's tendency to make poor choices?

There's no shortage of places to find this argument, but the obligatory link goes to Free Northerner for One More Condom In The Landfill, a particularly good presentation of the idea.

In a broad perspective the point is correct – empirically, men with more psychopathic traits, less agreeableness, and greater narcissism have more sexual partners.

On the other hand, it is kind of ironic that the pickup artist community – one of the few communities to be perfectly honest about the above point – has become obsessed with scoring the hottest girls and denigrating the others, no matter how perfect they might otherwise be.

The complaint tends to be "You women keep asking where the good men are, but they're right where you left them when you refused to date them because you only cared about cockiness and bulging muscles." The countercomplaint might be "You men keep asking where the good women are, but they're right where you left them when you refused to date them because you only cared about stylishness and big breasts."

I also suspect (though I have no evidence) that it is primarily the hotter women who have been socialized to be irrationally attracted to “bad boys”, and that pickup artists’ disproportionate focus on this demographic skews their assessment of the problem.

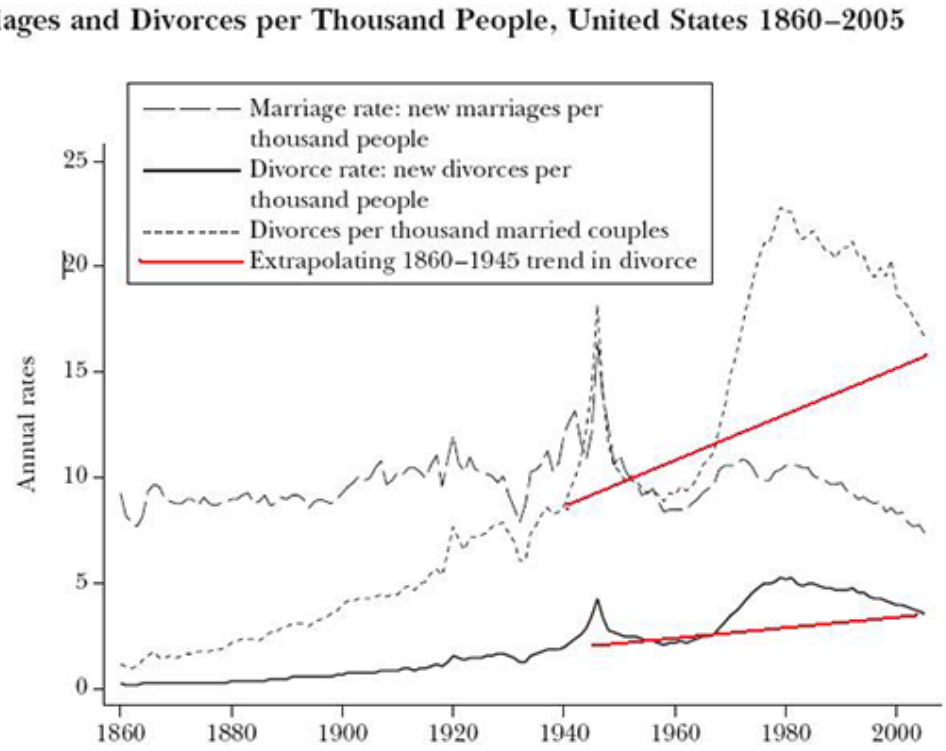
If one were to phrase the problem as “Men and women both make stupid and counterproductive sexual choices; how can we optimize for avoiding those?”, then that might make the sane 30%-or-so of feminists join the conversation and get something done.

If you phrase the problem as “Those women make stupid and counterproductive sexual choices, how can we shift the balance of power toward men?”, even the sane 30%-or-so of feminists will ignore and oppose you, and with good reason.

I have no idea how to solve the object-level problems, by the way, although I would tentatively recommend my own strategy of sidestepping the problems with both hot men and hot women by dating a hot genderqueer.

5.2: Are Progressive values responsible for rising divorce rates?

Let’s get the obvious objection out of the way first: divorce rates have been falling since about 1980. They’re now at their lowest level since 1970 or so, and dropping still.



The other thing this graph tells us is that rising divorce rates were a phenomenon very specific to the period about 1965 – 1975. This was a good decade for liberal values, but little moreso than decades before and after it. The strictly time-limited nature of the phenomenon suggests something more specific (and no, it's not no-fault divorce laws). The Pill, which came out in 1960, is an *extremely* plausible candidate, but a full treatment of this topic is beyond the scope of this essay.

Now that the obvious objection is out of the way, let's discuss some less obvious objections. If progressive values cause divorce, how come people with more progressive values are less likely to divorce? College-educated women have about half the divorce rate of the non-college-educated (source). More conservative states have higher divorce rates than more liberal states (source). Atheists have divorce rates below the national average (source). Some of these factors seem to remain even when controlling for wealth and the other usual confounders (source, source). The link between sluttiness and stable marriage mentioned above reinforces this point.

I think this data is consistent with the following theory: new technology and changing economic conditions produced a strain on family life that was reflected in an explosion in divorce rates. Society's memetic immune system sprung into action to contain the damage through the creation of new laws, institutions, and social norms. People who adopted the new ways survived the crisis and their family lives returned to a sort of normal. People who failed to adapt…well, don't be one of those people.

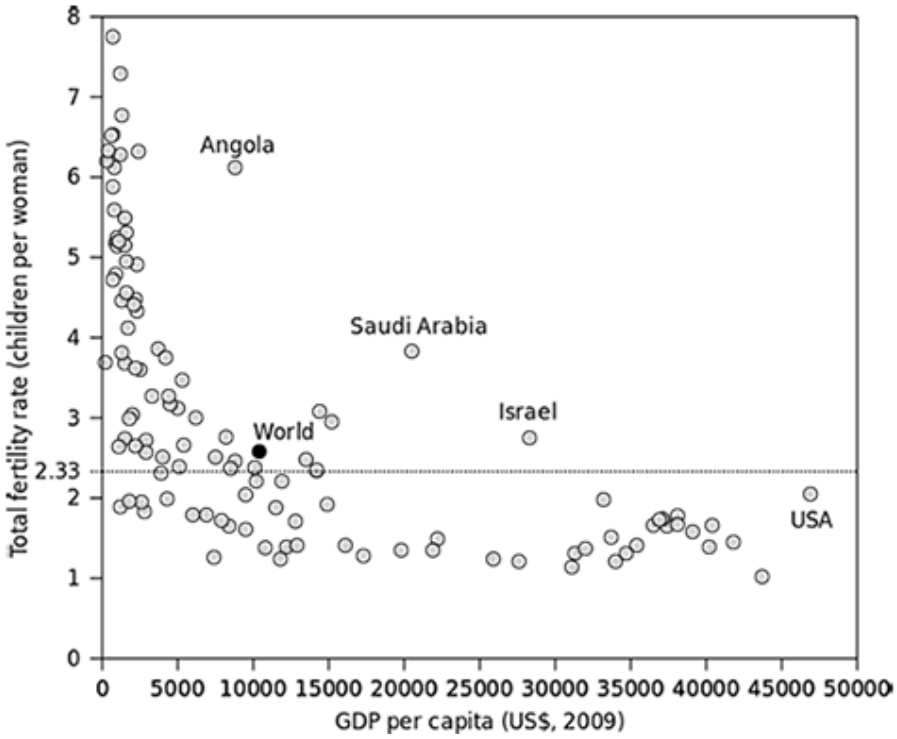
The new norms created by the memetic immune system are exactly the progressive values that Reactionaries blame for the damage: marrying later, trying more partners, using more contraception, having fewer children.

This theory explains both why the progressive values arise at the same time as the broken families, but also why people with progressive values are less likely to have broken families than others.

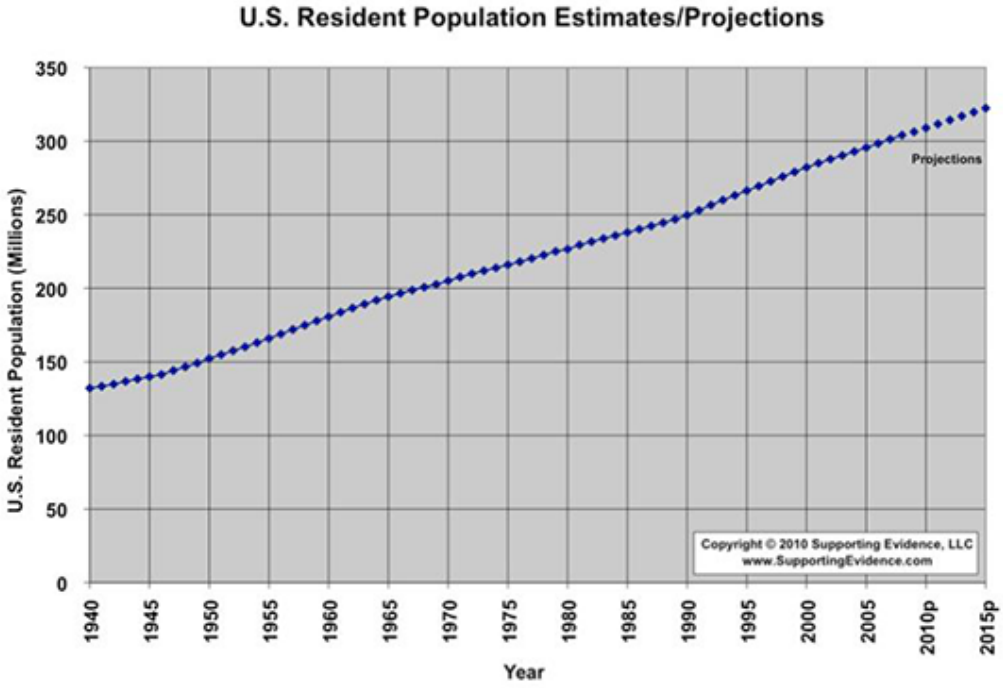
The data on illegitimate children and single motherhood mirror the data on divorce and do not require a separate discussion.

5.3: Are we headed for a demographic catastrophe?

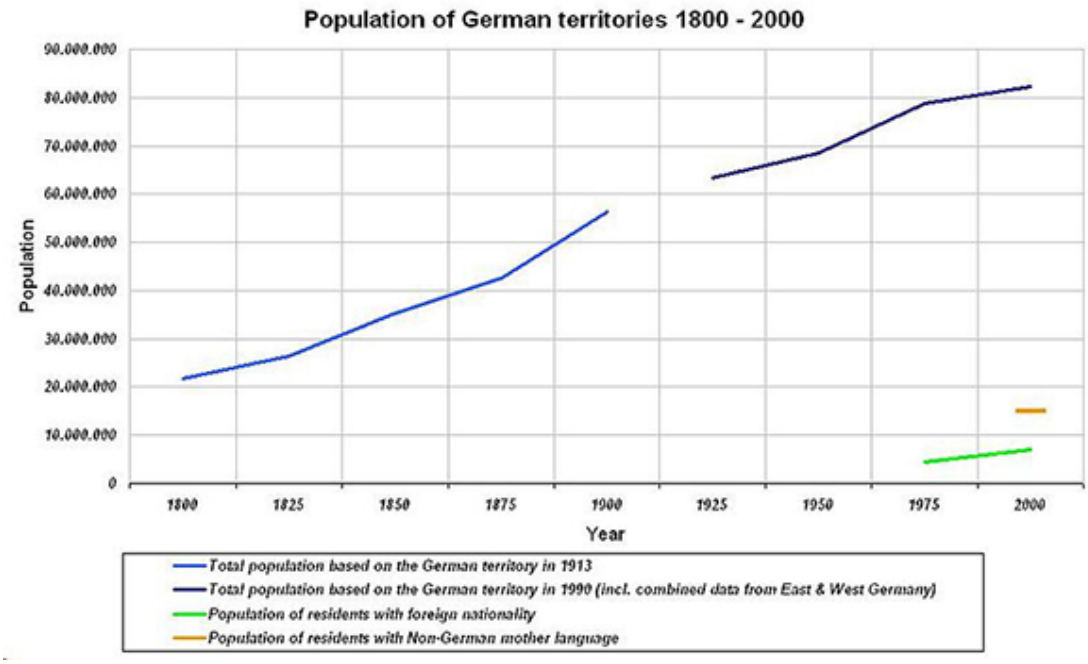
First of all, before we pretend that the minutiae of who has which values and who goes to church how many times affects fertility rate much, let's see the inevitable GDP/fertility rate graph:



And before we worry about the United States experiencing demographic collapse and tumbleweeds rolling through the streets of New York City, let's double-check to make sure that US population isn't a near-perfectly straight upward-trending line:



Western Europe?



A few countries do have demographic problems. Singapore, for example, has the lowest fertility rate in the world – 0.79, 224th out of 224 countries. It should probably do something about that. But given that it's generally accepted to be the most Reactionary country in the world, it's hard to blame this one on Progressivism or suggest Reactionary values as the answer.

5.3.1: But what if I am racist? Isn't it possible that fertile minorities and immigrants are hiding a fertility deficit among precious, precious, white people?

According to Edmonston et al's projection of US racial fertility trends:

> In 2100, the total U.S. population will eclipse 550 million people, and the racial composition of the country will be 38.8% white, 30.6% Hispanic, 15.6% black, 14.9% Asian and Pacific Islander, and 1% American Indian.

The absolute number of white people will be only a few million less than today, 209 million. That's more than enough to run a wide selection of excellent country clubs, or achieve whatever other strategic aims we need a large white population for.

Perhaps most gratifying if you are a racist, the percent of black people will increase only about three percentage points. The biggest increase will be in Asians, a so-called model minority.

After that? If there are still biological humans in organic bodies transmitting genes naturally much after 2100, we have *much* bigger problems than race on our hands.

5.3.2: Are we headed for an idiocracy?

Poor, uneducated, low-IQ people have higher fertility rates than wealthy, well-educated, high-IQ people in almost all countries. Therefore, one might worry that this will have a dysgenic effect, selecting against genes for intelligence until eventually everyone is stupid or has other undesirable quantities anticorrelated with wealth and education. This was the premise of the movie *Idiocracy*, and in principle people are far too quick to dismiss it.

But in practice, the effect is too small be significant. Richard Lynn, who is the closest we will get to an expert on dysgenics, calculates that American society as a whole is losing 0.9 IQ points per generation. So by 2100, people will have lost on average 4 IQ points.

Since it's hard to get a good intuitive graph of what 4 IQ points means, consider that IQ has been increasing by about 3 points per decade (average is still 100, but only because they recalibrate it). So absent any further Flynn Effect, losing 4 IQ points would take us back to…about as smart as we were back in 2000. I won't say that won't be unpleasant – the people of that era elected George W. Bush, after all – but it's not quite convert-all-written-language-to pictograms-because-everyone-has-forgotten-how-to-read level unpleasant.

And what comes after 2100 doesn't matter, because even on the off chance we're still using human brains to reason at that point, it sure won't be human brains in which the genes have been left to chance. To paraphrase Keynes, in the long run we're all either dead or cyborgs.

5.4: Aren't modern dogmas about race and sex and sexuality stupid and evil?

Let me be clear here. There is no excuse for the sort of extremist folk social justice crusades one can find on Tumblr or Twitter or Freethought Blogs. With a few treasured exceptions they are full of nasty and hateful people devoid of intellectual integrity and basic human kindness, and I am suitably embarrassed to be in the same 50%-or-so of the political spectrum.

Then again, there are lots of nasty and hateful conservatives and reactionaries devoid of intellectual integrity and basic human kindness too. Go take a look at Free Republic. Maybe we can call it a tie?

But this has surprisingly little bearing on the particular question above. As Christians are obligated by circumstance to point out, an idea is not responsible for the quality of people who hold it. And modern dogmas about race are agreed by very nearly everyone – including most Reactionaries! – including you! – to be both correct and very important.

Three hundred years ago, a pretty high percent of Americans were okay with black people getting kidnapped, enslaved, forced into back-breaking labor on plantations, raped, separated from their children, whipped if they protested, worked to a very early death, and then replaced with other black people.

Nowadays Reactionaries like to think of themselves as racist just because they believe the average black IQ is a standard deviation below the average white IQ. But one standard deviation implies that about a fifth of black people are smarter than the average white person. If you were to go back to 1800 and tell a conference of the most extreme radical abolitionists that you thought a fifth of black people were smarter than the average white person, they would laugh and not stop laughing until they died of laughter-induced asphyxiation.

And at least there the traditional and modern stereotype are still going the same direction. Did you know there used to be a stereotype that Jews were stupid and boorish and didn't belong in polite society? A stereotype that Chinese people were dumb? A stereotype that black people were bad at sports? To make a corny statistics pun, there seems to be very poor inter-hater reliability.

Homosexuality is little different. Reactionaries take a bold stand against sexually suggestive displays at gay pride parades or whatever, but when it comes to why two people who love each other can't get married because they're both the same gender, they tend to be just as confused as the rest of us. Mencius Moldbug writes:

> Although I am straight as an iron spear, I happen to see nothing at all wrong with "gay marriage." In fact I am completely sympathetic to the Universalist view, in which the fact that couples have to be of opposite sexes is a sort of bizarre holdover from the Middle Ages, like the ducking-stool or trial by fire. It's not clear to me why homosexuality, which obviously has some extremely concrete biological cause, is so common in modern Western populations, but it is what it is. However, because I am straight etc, and also because I'm not a Universalist, I happen to think the issue is not really one of the most pressing concerns facing humanity.

Moldbug is welcome to his opinion on what is or isn't one of the most pressing concerns facing humanity (I would have said a couple of brain-dead Internet thugs from Gawker beating up on a random Twitter celebrity isn't one of the most pressing concerns facing humanity, but to each his own) but I wonder if Moldbug notices that merely his unconcern on this issue makes him in let's say the 95th percentile of most Progressive Americans who have ever lived. 95% of Americans throughout history have been *quite* certain that eradicating sodomy *was* one of the most pressing concerns facing humanity, and boy did they act on that belief.

In fact, if we put a Reactionary in a time machine headed backward, and made it stop when the Reactionary was just as racist, sexist, et cetera as the US population average at the time, I predict they wouldn't make it much past the 1970s. Go into the 1960s and you get laws banning colleges from admitting both black and white students to the same campus (one helpfully specified that the black and white campuses could not be within twenty five miles of one another).

Now, there's no problem with this – except for Nixon and disco, the 1970s were no worse than any other period. But Reactionaries insist that all Progressivism since 1600 has been part of one vast and monstrous movement – maybe a religious cult, maybe a sinister power-play, maybe just the death throes of the western intellectual tradition – dedicated to being wrong about everything. And that a very big part of this vast movement focused on race. And when they have to whisper "Except we agree with 99% of what it did, right up until the past couple of decades, and in fact they got it right when everyone else was horribly, atrociously wrong", that is – or at least *should be* – kind of embarrassing.

5.4.1: But there's a clear difference between the past policies Reactionaries support and the modern ones they oppose. Past policies were going for equality of opportunity, modern ones for equality of results. Isn't seeking equality of results laden with too many assumptions?

Arguing about whether a post-racial society should provide equality of opportunity or equality of results is a little like arguing about whether in the worker's paradise, everyone should have a pony or everyone should have *two* ponies.

Right now, there is not even equality of opportunity. Rigorous well-controlled study after rigorous well-controlled study has shown that women and minorities face gigantic amounts of baseless discrimination in various areas, most notably employment. This remains true even when, for example, the experiment is sending

perfectly identical resumes out to companies but with the photo of a black or white guy at the top.

Once we have equality of opportunity, *then* we can start debating whether we should go further and try for equality of results. Until then, it's kind of a moot point.

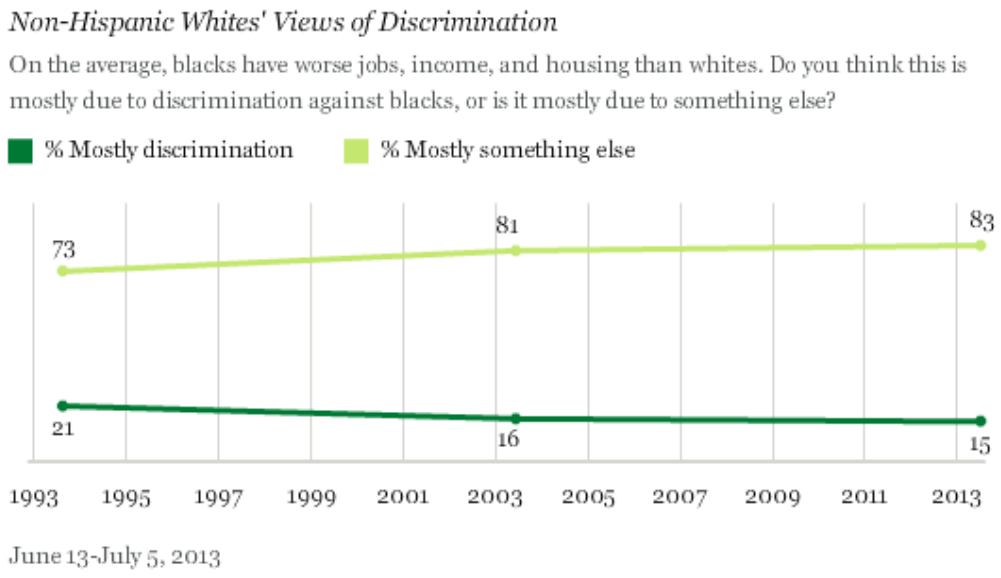
5.4.2: What about the studies that have shown black people have lower IQ/higher violence/other undesirable trait than white people?

If genetic differences across races prove real, this would be a good argument against seeking equality of results, but no argument at all against continuing to seek equality of opportunity – which, as mentioned above, mountains of rigorous well-controlled studies continue to show we don't have.

If, as the scientific racists suggest, black people have an average IQ of 85 compared to the white average of 100, then there is *still* a pretty big civil rights battle to be fought getting the average black person to do as well as the average white person with IQ 85. After controlling for IQ, the average black person is still twice as likely to be in poverty, 50% more likely to be unemployed, and 250% more likely to be in prison (source, other gaps appear to disappear or reverse once IQ is controlled; see link for a more complete analysis.)

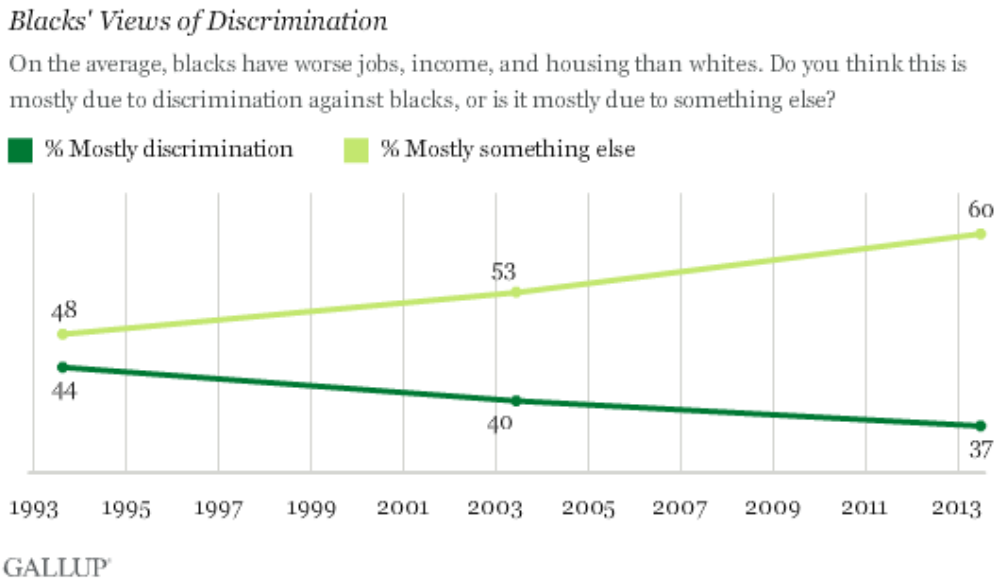
5.4.2.1: But this is exactly the kind of discussion progressives won't let us have! It is an unquestioned dogma of our society that all cross-racial differences must be based entirely on discrimination! In fact, people educated in public schools are incapable of even conceiving of the possibility that they could be otherwise! How are we supposed to be able to disentangle equality of opportunity from equality of results in such people?

From this Gallup poll:



83% of white people agree that the poor position of blacks in society is mostly not due to discrimination.

Want to see something even cooler?



60% of *black people* agree that the poor position of blacks in society is mostly not due to discrimination.

So no, doubting that all racial disparities in the US are due to discrimination isn't a thought crime. It's the majority position, even among black people themselves.

True, the number of people willing to consider genetic differences in particular would probably be far lower. But the great (and very legitimate) fear motivating more-than-academic interest in this question – that white people will forever be blamed for and forced to atone for minorities' problems – is one that can be talked about productively and perhaps banished.

5.4.3: Even if the establishment has not managed to completely ban all discussion of race that contradicts their own ideas, isn't it only a matter of time before political correctness takes over completely?

It's hard to measure the power of the more intellectually bankrupt wing of the social justice movement, but as best I can tell it does not seem to be getting more powerful.

According to Rasmussen, support for "political correctness" is declining in America. As we saw above, fewer and fewer people are willing to attribute black-white disparities to "racism" over time. Gallup finds that in the past decade, the percent of blacks satisfied with the way blacks are treated has gone up nearly 10% (I can't find similar numbers for white people, but I bet they're similar). Both white and black people are about 25% less likely to consider the justice system racially biased than 20 years ago. The percent of whites who think government should play "a major role" in helping minorities has dropped by 10 percent since 2004; for blacks, there is a similar drop of 14 percent.

The percent of people who think women have equal job opportunities to men has gone up 15% in the past nine years. Women are less likely to identify as feminists than twenty years ago, and support for affirmative action is at historic lows.

Here we see really the most encouraging combination of trends possible: actual racism, perceptions of racism, and concern about racism are all decreasing at the same time.

5.4.3.2: So how come social justice people have been making so much more noise lately?

My guess is changes in the media. The Internet allows small groups to form isolated bubbles and then fester away from the rest of society, becoming more and more extremist and paranoid and certain of themselves as their members feed upon each other in a vicious cycle.

Of course, as Reactionaries, you wouldn't *possibly* know anything about that.

At the same time, the relative anonymity of the Internet promotes bad manners and flame wars and general trollishness. It's not just that the writer is anonymous and therefore doesn't fear punishment for what he or she says. It's that their *enemy* is some nameless evil, rather than a person with a face whom they will treat as a human being.

And again at the same time, the national media has become more and more efficient at detecting outrageous events associated with some small town or some B-list celebrity and publicizing them to the entire world. This allows the hatred of the entire world to be focused on a single random person for a short period of time, which usually results in that person's life being ruined in a way that would be impossible without this media efficiency.

But these processes are at least partly nonpartisan. With a rise in extremist online social justice has also come a rise in groups that didn't even exist before, like men's rights advocates.

5.4.3.2.1: Still, isn't the fact that progressivism was responsible for this sort of zealous and hateful social justice movement is a point against it?

I identify the worst parts of the social justice movement as basically reactionary in their outlook, even though from a coalition politics point of view they have been forced to ally with progressives.

Chief in this assessment is their strong beliefs that some topics should be taboo and bowdlerized from society. In the old days, you

would ban books because they talked too much about sex. In the new days, we laugh at their prudishness, but still seriously debate banning books because they are "demeaning towards women" or "trivialize rape culture". The desire to ban books that promote different sexual norms than we ourselves promote hasn't changed, only the particular sexual norms we are enforcing.

The same is true of race. In the old days, we would ban books that insulted the King or the upper classes. In the new days, we ban books that insult the poor, or disprivileged or disadvantaged classes. Again, the desire to ban books insulting the classes we like doesn't change, only to which classes we afford this privilege.

Real Progressivism is Enlightenment values – like the belief that free flow of information is more important than any particular person's desire to "cleanse" society of "unsavory" ideas. Real Reaction is the belief that free expression isn't as important as making sure people have "the right" values. Upper-class white Reactionaries will try to enforce values protecting upper-class white people. Lower-class minority Reactionaries will try to enforce values protecting lower-class minorities. Whatever. They're still Reactionary.

Likewise, real Progressivism is color-blind. It may be *sophisticatedly* color-blind, which involves realizing that just saying "I'm going to be color-blind now, okay?" doesn't work, and that affirmative-action type policies may paradoxically lead to more genuinely color-blind results. But it would be unlikely to promote the idea that people should have racial pride, or that one particular race is evil and is not allowed to have racial pride. "White people should identify strongly with white culture; black people have no culture" is the upper-class white Reactionary slogan. "Black people should identify strongly with black culture; white people have no culture" is the lower-class minority Reactionary slogan. "Lots of races have culture but let's ignore them and let individuals identify with what they personally like" is

the academically-neglected but still-popular true Progressive position.

Finally, real Progressivism opposes segregation in all its forms. Upper-class white Reaction says that it's necessary to protect white people from being "polluted" by black culture like rap music. Lower-class minority Reaction says that it's necessary to stop white people from "appropriating" black culture like rap music. Either way, we get white people not allowed to listen to rap music. Progressivism is the position contrary to both: that everyone can listen to whatever music they damn well please.

The conservative nature of social justice isn't surprising if you, like me, believe the liberal/conservative divide mirrors a self-expression/survival divide – more simply, whether or not you feel safe. As society becomes more economically and politically secure, we expect it to become more liberal and progressive. But we also expect the subgroups of society that are least secure to remain conservative, and to continue to use conservative strategies to protect themselves in their unsafe environment. Those subgroups are women and minorities.

Because more liberal white people are more likely to be tolerant toward minorities and the poor, minorities and the poor are by political necessity forced to ally with liberal parties. But when we are able to separate issues out from political coalition-building and self-interest, the natural tendency of economically and physically insecure minorities to be more socially conservative shows itself. Black people are more religious, more likely to support amendments banning gay marriage, and more likely to oppose stem cell research, abortion, and out of wedlock births.

If you do not like certain extreme versions of social justice, then fighting their Reactionary memes favoring poor minorities with your own Reactionary memes favoring rich whites is unlikely to work. At best you would just end up with two angry clans demanding more power for them personally; more likely financial

and signaling incentives will prevent rich whites from wanting to take their own side in a conflict and everyone will just ignore you. A better strategy would be to take the moral high ground and promote Progressive memes to both sides.

5.5: Is our society hopelessly biased in favor of minorities and prejudiced against white people?

The most visible parts of society, like affirmative action and conversational norms around political correctness, are biased in favor of minorities and against white people. But this is intended to counter less visible parts of society, which are biased in favor of white people and against minorities. Whether this gambit works is anyone's guess. See An analysis of the formalist account of power structures in democratic societies for a more careful evaluation of this claim.

5.6: One particularly annoying politically correct idea is the demand that everyone feel guilty about colonialism. Colonialism helped industrialize the developing world. Wasn't the Progressive attempt to "help" the developing world through enforced decolonization and self-rule actually a big step backwards?

There are a couple of studies on this question, but all have their issues. A particular problem in the comparison of colonized to uncolonized countries is the possibility that more prosperous countries would be more likely to attract colonization *and* more likely to successfully resist potential colonizers. This makes an attempt to formally compare colonized with never-colonized countries directly nearly impossible.

I am least dissatisfied with Sylwester 2005, which compares colonial countries before, during, and after decolonization. It finds that:

> There was no decrease in growth [for newly independent countries] relative to the alternative of remaining a colony.

> The reason why decolonizers exhibited lower growth than did those not concurrently undergoing a political change is that decolonizers grew slower than did nascent countries. These results provide evidence against the claim that this type of political transition caused lower growth than experienced previously. There is no evidence of transitional costs.
>
> The paper also finds that previously independent countries grew faster than did the existing colonies. Whether or not a region is independent or controlled by an external power appears important for growth outcomes"

In other words, countries grew faster after independence than they did as a colony. This provides some support for the leftist idea that colonial powers drained more resources than they introduced, at least towards the end of the colonial age.

5.6.1: Forget economics, then. Wasn't decolonization a human rights disaster, considering all the civil wars and coups and mismanagement in former colonies that could have been prevented by a competent colonial government?

Everyone from every side of the political spectrum agrees decolonization could have been handled better. It might be that no decolonization at all would have been better than decolonization the way the Great Powers historically went about it. And it's hard to excuse all the civil wars and mismanagement that caused.

On the other hand, the colonial era wasn't exactly free of bloody wars either. Colonial wars included the Mahdist War (100,000 deaths), the Algerian Revolution (500,000 – 1.5 million deaths), the Rif War (70,000 deaths), the Italian-Ethiopian War (500,000 deaths), the Mau Mau Rebellion (20,000 deaths), Mozambique War Of Independence (80,000 deaths), Angolan War of Independence (50,000 deaths), the Herero Genocide (100,000 deaths), the Java Wars (200,000 deaths), Sepoy Mutiny (~100,000 deaths), the Mad Mullah Jihad (100,000 deaths, but on the brighter side, an awesome name) Philippine-American War (220,000 deaths), First Indochina

War (200,000 deaths), Aceh War (100,000 deaths) et cetera, et cetera, et cetera.

If we don't limit ourselves to just wars, and include famines, genocides, and general mismanagement, we can add Congo Free State (8 million deaths), genocide of Brazilian Indians (?200,000 deaths), forced labor in Portuguese colonies (250,000 deaths), forced labor in French colonies (200,000 deaths), Italian colonial genocide in Libya (125,000 deaths), French colonization of Algeria (500,000 deaths, European eradication of Native Americans (350,000 deaths), and the Australian and New Zealander eradication of aborigines and Maori (440,000 deaths). If we are willing to count famines worsened by colonial mismanagement we can go almost arbitrarily high, 20 million deaths or more.

It is certainly possible to *imagine* a wise and paternalistic colonial government coming in, cleaning up after native misrule, and introducing things like sanitation and industrialization. But that's not what happened. It's not fair to compare an imaginary ideal version of one policy with the real-world version of another.

5.6.1.1: Weren't a lot of those colonial wars and human rights abuses actually caused by demotism and Progressivism? If people hadn't revolted against their colonial masters, there wouldn't have been these bloody colonial revolts.

Not a straw man!

The first answer is that even if we accept this weird premise, there are still hundreds of colonial atrocities that do not stand excused. Many of the above conflicts occurred during original colonial invasions, and a tendency to resist those hardly requires demotism. Others were simple genocides, during which resistance was minimal.

But let's not accept the premise. I admit placing blame is complicated. To give just one example, thousands of homosexuals were killed in Nazi Germany. We usually blame the Nazis for this. But from a formal math point of view, it would be equally valid to

blame homosexuality. After all, if not for homosexuality, those people would not have been killed, Nazis or no.

How to avoid such bizarre conclusions? One method is moral – even if both Nazism and homosexuality were to blame according to purely mathematical casual models, Nazism seems more *morally* to blame. Another method is practical- homosexuality is as old as the human race and probably not going away, so it's easier to view homosexuality as a constant and vary Nazism than it is to hold Nazism as constant and vary homosexuality.

We can apply these same methods to the colonial wars. Morally, the colonized people seemed to be morally in the right – they were sitting around trying to live their ordinary lives when people invaded and tried to turn them into forced laborers. And practically, the desire for self-rule is older and harder to root out than the colonialism. Indeed, colonialism pretty much died off after a century or two, and the desire for self-rule is stronger than ever.

Some Reactionaries would contest this hypothesis. They would say that it is only the spread of Progressive ideas that make people want to revolt against their colonial masters – that if not for the *New York Times* deliberately sowing pre-revolt memes, no one would consider this a worthwhile thing to try.

Historical counterexamples abound, but the Jewish-Roman Wars (66-135 AD) seem like a particularly good one. If they don't appeal to you for some reason, pick your own favorite example out of Wikipedia's List of revolutions and rebellions.

And as we saw above, if Progressivism is an inevitable historical reaction to rising technology and security, rather than a meme spread by the New York Times or anyone else, then saying "My scheme would have worked if not for the spread of Progressive ideas" is no more virtuous than saying "My scheme would have worked if not for the conservation of matter". Congratulations, you've found something that might have been a good idea in an alternate universe that ran on different rules.

5.6.2: Even if colonialism was historically bloody, wouldn't today's human-rights-obsessed, racism-hating era be able to sustain a type of colonialism that gives the good parts without the evil?

Yes, it's possible that modern progressive ideals would be able to rescue colonialism. But it's hard to imagine a nation being simultaneously progressive enough to colonize other countries wisely, but still so unprogressive that it would want to. It would have to be a country whose progressivism evolved on a path much different to our own.

5.7: Are schools are places where children get brainwashed into leftist and blame-America-first values? Are all parts of history that don't fit with a progressive worldview whitewashed from the curriculum?

Our source here is James Donald, who for example says:

> History gets radically rewritten at ever shorter intervals, and all older history books are effectively banned. Consider, for example the ever more radical rewrites of the career of Daniel Boone, which ended with him being expelled from history altogether, and that today's student has no idea what "The shores of Tripoli" refers to. Ninety nine percent of what students used to be taught not very long ago, is now unthinkably controversial, shocking, and disturbing…
>
> Look [these things] up in a history book written before the days of hate-America-first history. The New Century Speaker for School and College, published 1905.
>
> Of course this would require you to read old books, but old books are like kryptonite to a progressive. Since they were written by dead white males, no respectable person will read them for fear that dangerous and forbidden thoughts might contaminate his brain. Like a vampire confronted with a bible, a progressive will cringe in fear before any dangerously old

> book. Ever since 1905 or so, kids have been taught hate-america-first history.

I worry James is confusing the sign of a value with the sign of its derivative. Certainly schools are becoming *more* willing to discuss leftist issues. But are they now *disproportionately* willing to discuss them?

Let's take the example of Columbus. Modern Americans are taught not only the old history that Columbus was a brave explorer who sailed forth to boldly discover that the Earth was round, but also the new history of "yeah, but he was bad for the Indians". The feeling I got was that sure, Columbus was all nice and well, but his bold voyages paved the way for later people to settle the New World which sort of by coincidence hurt the Indians because people were squatting on their ancestral lands. This is about as far as so-called liberal schools will go, and this is probably the sort of progressivism being introduced to history classes which James is complaining about.

But actually, Columbus was…well, The Oatmeal is kind of a low-status source to link to, but I think they said this one better than I could. It starts off with :

> Upon his arrival, he demanded that the Lucayan [Indians] give his men food and gold, and allow him to have sex with their women. When the Lucayans refused, Columbus responded by ordering that their ears and noses be cut off, so that the now disfigured offenders could return to their villages and serve as a warning to others. Eventually, the natives rebelled. Columbus saw this as a perfect excuse to go to war, and with heavily armed troops and advanced weaponry, it wound up being a very short war. The natives were quickly slaughtered…there are eyewitness accounts of fallen Lucayan warriors being fed to hunting dogs while they were still alive, screaming and wailing in agony as the dogs feasted on their limbs and entrails.

(a commenter points out that some of its other claims are exaggerated)

As much as James may complain about how people vaguely mutter about something something Indians something on Columbus Day, I bet he didn't learn this in school. In fact despite his protestations, I bet he didn't learn very much leftist history at all in school, given that he thought Eugene V. Debs was a Supreme Court case.

One day, our school curriculum may become so leftist that the Right needs a book like A People's History of the United States or Lies My Teacher Told Me (which was created not by armchair contemplation of what society's biases *must* be, but by reading twelve actual history textbooks and spotting the actual lies in them). But that day hasn't come yet.

What is James' own evidence for a leftist bias? As far as I can tell, they're things like that US classrooms keep going on about US enslavement of black people, but never mention the (African) Barbary Pirates enslaving white Americans. But this may have less to do with liberal bias and more to do with the fact that, as far as I can tell, only 115 white Americans were ever enslaved by the Barbary Pirates (and then released a few years later), whereas about 500,000 African slaves were brought to America, kept in slavery for centuries, precipitated the bloodiest war in our country's history, and then became a racial group that makes up 12% of Americans today – over forty million people.

Oh, and actually, I *did* learn about the Barbary Pirates in history class, thank you very much. So it seems that prediction of James' has been disconfirmed. Although he seems to have thought the government shutdown might end with Tea Party members and lawmakers being shipped to concentration camps, so I imagine having his predictions disconfirmed is a pretty common occurrence for him.

I apologize for the insulting tone of this FAQ entry, but I was accused of cringing in fear before old books, and being vampire-to-

Bible-level afraid to study history. That hurts.

6: Any last thoughts?

6.1: Does this mean you hate Reactionary ideas and think they have nothing to teach you?

Absolutely not. Compare to communism. The people who called themselves communists had some great ideas, like shorter workweeks and racial equality. It was just that the narrative they used as a framework for that idea – historical dialectic, workers controlling the means of production, violent revolution, destruction of capitalism, destruction of democracy – were horrible. Their ability to notice problems tended to be better than their specific policy proposals which in turn tended to be better than their flights of fancy.

I feel the same way about Reaction. Some Reactionaries are saying things about society that need to be said. A few even have good policy proposals. But couching them in a narrative that talks about the wonders of feudalism and the evils of the Cathedral and how we should replace democracy with an absolute monarch just discredits them entirely.

6.1.1: What *exactly* do you like about Reaction?

I like that they're honestly utopian. Their scathing attacks on everyone else for being utopian merely punctuate the fact, like the fire-and-brimstone preacher denouncing homosexuality whom everyone knows is secretly gay. The Reactionaries wants to throw out the extremely carefully fine-tuned machinery of modern society which evolved over several hundred years, and replace it with a bizarre Frankenstein's Monster of modern and traditional elements that they dreamed up in an armchair, which has never been tried before and which, they say, will instantly fix all social ills like crime and poverty and war.

And this is awesome. Utopianism – trying to think up amazing political systems that lie outside the local Overton Window – is

very nearly a dead art. The failure of the Communists' utopian designs probably killed it – the Right made "utopianism" into a dirty word so they could use it to bludgeon the Left, and the Left turned against utopianism en masse to avoid getting bludgeoned. Right now the only two permissible dreams of a better future are a society much like our own but a little more libertarian, or a society much like our own but a little more progressive. Boring!

The more utopian ideas we have the more sources we have to draw from when trying to decide which direction our own society should go in, and the broader the discourse becomes. Reactionaries are geniuses at inventing new systems that have never been tried before and some of whose components deserve serious contemplation. And if there was a science fiction book set in Moldbug's Patchwork or Royal California, I would buy it.

6.1.1.1: But?

There are a few good things you can do with utopianism.

You can use it as a generator for ideas that become gradually adopted into the mainstream, as mentioned above. Communism was good at this – in the US, instead of starting a revolution, they just helped spark the modern labor movement, which eventually came to coexist with the rest of the economy and is now probably a useful part of the memetic ecosystem.

You can use it to start interesting intentional communities. There were a couple of communist communes within capitalist countries; some people even built phalansteries, and more modern versions like Twin Oaks are more successful. You can start a non-communal subculture, like the polyamory movement. If you happen to have a free land, you start a country or subnational government – it worked for the early American settlers, and it may yet work for seasteaders. The Free State Project is another noble goal along these lines.

But until it works in an intentional community or something, trying to push it on everyone else seems premature and irresponsible.

6.1.2: If we don't do Reaction, does that mean we're stuck with a boring inoffensive centrist democracy forever and ever?

No. There are lots of extremely creative ideas for radical new forms of government that don't involve any Reactionary ideas at all. The better ones are off of the right-left spectrum entirely. Futarchy is my favorite. Or we could all just go live in the Shining Garden of Kai-Raikoth.

6.2.1: Has anyone written a response or rebuttal to this FAQ?

Ohhhhhh yes.

I am indebted to Reactionary blogger Legionnaire for putting together a good list of responses to this document, which I am reproducing here with only minor aesthetic changes.

RESPONSES TO PART 1: IS EVERYTHING GETTING WORSE?

Foseti – An Anti-Reaction FAQ

Xenosystems – The Decline Frame

Jim – Anti-Anti-Reactionary FAQ Part 2: Crime

More Right (Michael Anissimov) – Response to Anti-Reactionary FAQ, Lightning Round, Part 1

RESPONSES TO PART 2: ARE TRADITIONAL MONARCHIES BETTER PLACES TO LIVE?

Jim – Anti-Anti-Reactionary FAQ Part 1: Terror And Mass Murder (this limited its complaint to a single example and seemed quite fair, so I have since removed that example from this document)

Jim – Anti-Anti-Reactionary FAQ Part 3: Freedom And Monarchy

More Right (Michael Anissimov) – Response To Anti-Reactionary FAQ Part 2: Austrian Edition

RESPONSES TO PART 3: WHAT IS PROGRESS?

Jim – Progress

Jim – Anti-Anti-Reactionary FAQ Part 4: Ever Leftwards Movement

Anarcho-Papist – The Theory Of Demotist Singularity

Habitable Worlds – The Motives Of Social Policy

RESPONSES TO PART 4: SHOULD A COUNTRY BE RULED AS A JOINT-STOCK CORPORATION?

Anarcho-Papist – The Informal Systems Critique of Formalism

RESPONSES TO PART 5: ARE MODERN IDEAS ABOUT RACE AND GENDER WRONG-HEADED AND DANGEROUS?

Anarcho-Papist – On The Opposition To Sluttiness, Among Other Things

Free Northerner – Sex: A Response To Scott Alexander

Jim – The Anti-Anti-Reactionary FAQ: Sluts

MISCELLANEOUS RESPONSES

Nick Steves – Shots Across The Bow

Suntzuanime – Comment On Anti-Reactionary FAQ

I've only managed to read about 50% of these so far, but of the ones I have read, I am especially impressed with Anissimov's Lightning Round Part 1 and Free Northerner's post on sex issues as well-argued and pretty comprehensive critiques.

I will continue to update based on his list as a definitive resource, but if you've written something and want on here, post in the comments of this thread or email me and I will *eventually* get you up. This is likely to update very irregularly.

The Poor You Will Always Have With You

I'm gradually reading through responses to the Anti-Reactionary FAQ, but I'll take a moment to respond to this excellent and well-argued post from Habitable Worlds in particular because it points out an especially deep disagreement.

Scharlach from Habitable Worlds objects to my point 3.1.1, which claims that progressive ideals aren't particularly novel or modern because classical Rome shared many of the policies we most associate with progressivism. I mention welfare, strikes agitating greater rights for the poor, multiculturalism, religious syncretism, sexual libertinism, and utopianism.

Scharlach disagrees. He first points out that classical Roman "strikes" were not about greater rights for the "poor", per se, but for plebians, a class of non-nobles that actually included some very wealthy people. I accept his clarification, but I would add that modern progressive movements are happy to conflate "class made up of disproportionately poor people" with "poor people" as well, whether we are talking about the unemployed, inner city youth, minorities, high school dropouts, inhabitants of the Third World, or whatever. Heck, modern progressivism calls women a "minority" even though they make up 51% of the population just because it is a convenient way to lampshade their less privileged status. So I don't think it's especially unprogressive that "more rights for plebians" was the classical Roman rallying cry, rather than "more rights for the poor" per se.

But the crux of his objection is more philosophical:

> But the question is: do these seemingly "progressive" policies stem from what today we would consider progressivism? Do they have anything to do with "social justice"? We should remember that when looking back at history, curious similarities arise, but they do so at incongruous joints, and their existence may not signify anything but the fact that large-scale

political ecologies have limited practical expressions. Think of it this way: A society whose political discourse and ideals sanction welfare to the poor because it is believed that the underclass is genetically inferior, incapable of taking care of itself, and might revolt if not given enough food … that's a very different society from one whose political ideals sanction welfare because it is believed the poor have a right to good living standards or that the poor deserve welfare because it re-distributes goods rightly theirs but taken from them through an oppressive economic system.

Contemporary progressive policies emerge from ideals and discourses about morality, justice, oppression, and rights. The poor (especially the dark-skinned poor) deserve the welfare they get; it is theirs by Constitutional right. It is a moral and political imperative not to take away the welfare they receive and to give them more if possible. Progressives actively try to alleviate the shame once associated with receiving welfare. Pointing out that the poor in America have it pretty good is a distinctly right-wing thing to do. "Food stamps" are now "EBT cards" that look and function like debit cards. Medicaid patients sit in the same waiting rooms as patients paying high insurance premiums, and you can't tell the difference. (Well, you can, but …) Welfare in America has become a right, a moral imperative, a matter of justice and just desserts, a thing that brings no shame, a thing to be proud of, a thing to demand, a thing to stand up for…

So Scott Alexander is correct that social policies in ancient Rome look similar to contemporary progressive welfare policies. But were the motives the same? Did the poor and the plebians get free or reduced-cost corn, grain, wine, and olive oil … . because they deserved it? because it was theirs by moral and legal right? because it was a matter of social justice?

I'm not a classicist, so I'm willing to be corrected on this, but as near as I can tell, the Roman dole was wrapped up in

> discourses about a) the might and wealth of Rome and b) goddess worship. Welfare policies in ancient Rome were built upon very different ideals and emerged from very different motives than contemporary progressivism's welfare policies. Nowhere have I been able to fine a discussion of the Roman congiarium in terms of rights or justice. The dole was there because it made the emperor more popular and demonstrated the wealth of Rome to the people. What's more, the dole was personified as Annona, a goddess to be worshiped and thanked. Scott Alexander even recognizes this difference in motive when he says that ancient Romans "worshiped a goddess of food stamps."
>
> Indeed they did. And that's the whole point. When was the last time you heard welfare policies discussed in terms of worshipful gratitude, mercy, and thankfulness? If that were the discourse surrounding welfare policy, America would be a very different country. It seems that Roman welfare and American welfare are as different from one another as Jubilee is from abolitionism.

I will agree that the Romans used different philosophical justifications for their welfare state than do moderns, but before discussing this, a lengthy and kind of pointless also-not-a-classicist digression on why the difference may not be as big as Scharlach suggests.

If the essay is trying to compare the grateful Roman poor and the entitled, demanding modern poor, I propose that the Roman recipients of the *annona* were as entitled and demanding as any modern. Ancient Roman leaders automatically assumed any hiccup in the flow of free grain would lead to riots, and their assumption was justified. You may for example read the section on Roman food riots here. Particular high points are the riots of 22 BC, during which rioters threatened to burn the senators alive if they didn't produce enough free grain, and the riots of 190 AD, when Papirius Dionysius, the prefect in charge of the grain supply, accused political

enemy Marcus Aurelius Cleander of threatening it – the disturbance ended when the Emperor Commodus killed Cleander and his son and threw their heads out to the angry mob (which instantly calmed down and dispersed).

Or the essay may be trying to compare a Roman attitude of giving small strategic grants of welfare to the worthy with a modern attitude that everyone deserves as much welfare as they want at all times regardless of situation or else their human rights are violated. But here, too, I do not think the distinction is as great as is claimed. 83% of Americans believe people on welfare should be required to work, and only 7% oppose such a requirement. 69% believe that there are too many people on welfare and the criteria need to be stricter, compared to only 24% who believe the opposite. People who want welfare benefits need to jump through various bureaucratic hoops (some of which are actually kind of stupid) and usually receive them only for a limited amount of time.

(this interpretation would remind me of my frequent complaint that some reactionaries say "X is an unquestionable dogma of our modern society" when they mean "I heard about a college professor who believes X".)

So much for our pointless digression. Scharlach probably means something more like "Ancient Rome didn't have modern concepts of human rights and social justice." I agree with this. I just don't think it matters.

I assume Scharlach read my FAQ part 3.3, where I claim that progressive values are closely linked to urbanization and technological/economic growth. But he may not have read my We Wrestle Not With Flesh And Blood…, so I'm worried he might have interpreted me in 3.3 as claiming something like:

Urbanization + Growth -> Progressive Values -> Social Change

If that had been my thesis, then it would indeed be relevant that the ancient Romans didn't have our version of progressive values. Their social change would be a coincidence, unrelated to ours since it

missed the crucial middle step that determined the shape our social change would take.

But I'm not proposing that model. I'm proposing one that looks more like this:

Urbanization + Growth -> Social Change -> Progressive Values

(really the "social change" node should be called "pressure for social change", and it and the "progressive values" node should have little circular arrows both pointing at each other, but let's keep it simple)

Let me give an example of what I mean.

A 25th century historian, looking back at our own age, might notice two things. She would notice that suddenly, around the end of the 20th century, everyone started getting very fat. And she would notice that suddenly, around the end of the 20th century, the "fat acceptance movement" started to become significant. She might conclude, very rationally, that some people started a fat acceptance movement, it was successful, and so everyone became very fat.

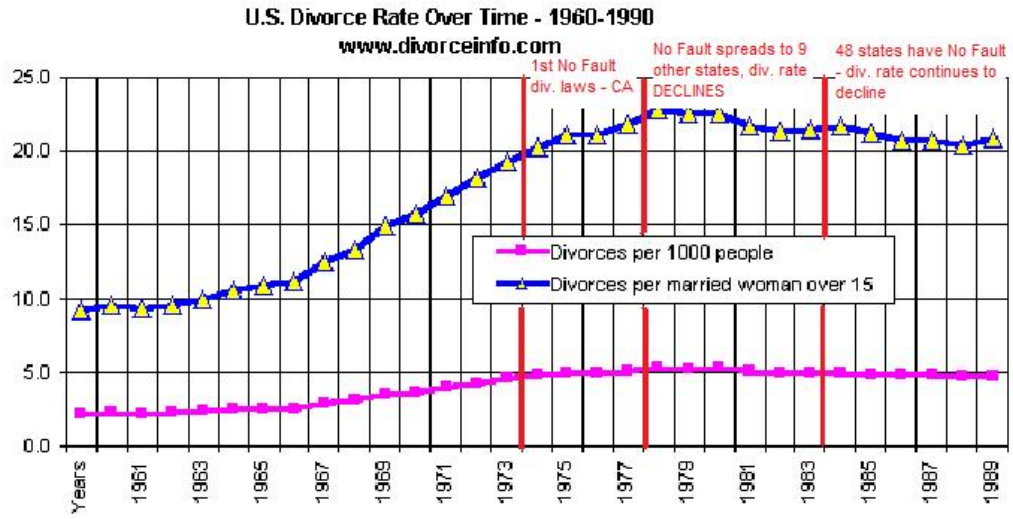
With clearer knowledge of our era, we know better. We know that people started getting fat for, uh, reasons. It seems to have a lot to do with the greater availability and better taste of fatty, sugary foods. It might also have to do with complicated biological reasons like hormone disrupters in our plastics. But we have excellent evidence it's not because of the fat acceptance movement, which started long after obesity rates began to increase. If we really needed to prove it, we could investigate whether obesity is more common in populations with good access to fat acceptance memes (like, uh, Wal-Mart shoppers and American Samoans).

To us early-21st century-ites, it's pretty clear why the fat acceptance movement started now. Its natural demographic is fat people, there are more fat people around to support it, they feel like they have strength in numbers. and non-fat people are having trouble stigmatizing fat people because it's much harder to stigmatize a large group than a small group (no pun intended).

Does this have any relevance for the sort of thing reactionaries talk about? Yes. Let's look at divorce.

From a historical perspective, no-fault divorce was legalized in the early 1970s, and divorce rates were skyrocketing in the early 1970s. It is *incredibly* tempting to want to attribute skyrocketing divorce rates to easy-access no-fault divorce.

It's also wrong. From an excellent article I entirely recommend:



Just from the graph it should be clear how little no-fault divorce mattered, but if you need more formal research it has certainly been done. Even the conservative Institute For Marriage and Public Policy admits in its review article on the subject that "divorce law is not the major cause of the increase in divorce over the last fifty years", and that even the small bump from no-fault provisions "while sustained for a number of years, eventually fades and the divorce rate moves back to trend".

I'd guess that the explanation for why skyrocketing divorce rates and no-fault divorce both happened in the early 70s is a lot like the explanation for why skyrocketing obesity rates and fat acceptance both happened in the early 2000s. Lots of people started getting divorced. Under older, stricter divorce laws, this required couples who wanted divorces to manufacture some bogus complaint with the help of lawyers, an embarrassing and expensive process. Eventually the number of people divorcing or wanting to divorce became

sufficiently large to form a good political lobby, and the people not involved in the divorce process couldn't keep stigmatizing divorcees because there were too many of them for it to be easy or convenient. So the divorce lobby won and passed no-fault divorce laws.

I don't deny that sometimes these ideological movements and the laws they pass have some effect, like the small, quickly fading effect of divorce laws mentioned in the quote above. That's why I wanted little circular arrows between "Social Change" and "Progressive Values" above. I'm just saying these effects are small and not particularly interesting. They're the tail wagging the dog.

And I don't deny that the progressive movement pushing a social change often exists before the social change does. If 100 years from now the existence of vat-grown meat causes all factory farming to shut down, no doubt PETA will claim victory. But just because PETA pushed for the event, and then the event happened, doesn't mean PETA was the main cause. At best, they kept pushing but it was only the technological change that helped them gain power and respect and enact their positions. At worst, if they didn't exist then within ten minutes of the invention of vat-grown meat some other group would have sprung up to accept the easy moral victory it provided.

So let's get back to Rome.

Scharlach points out that the value system associated with Roman welfare was different from the value system associated with our own welfare system.

Ancient Rome had a population of about a million people crowded together, a government vulnerable to the mob, and resources to spare. I propose those situations will, more often than not, inspire a welfare system. They did it in ancient Rome, and they're doing it in modern DC.

According to legend, Frederick the Great declared of his conquests: "I will begin by taking. I shall find scholars later to demonstrate my perfect right" (okay, Reactionaries, I will admit Frederick the Great

was hella cool). If Frederick was in the welfare business, he might have said "I will begin by giving welfare. Later, I will find scholars to come up with a philosophy supporting welfare."

And just as any historical account of why Frederick conquered new territories should focus on his self-interested goals rather than on whatever justifications his scholars later cooked up, so an account of why we give welfare should focus on the economic, material, and technological conditions that inspire it, rather than fretting over how one society talked about the goddess Annona and another talked about social justice. I'm sure if Frederick conquered both classical Rome and 21st-century America, his Roman supporters would declare he was following the will of Jupiter, and his American supporters would declare he was trying to help disprivileged minorities. It would be the historian's job to see through that (and also to sort out what I expect would be a very confusing timeline of Frederick's life).

Which brings us back to Rome one last time. I didn't discuss the Roman welfare state in isolation. I mentioned it in the context of Rome being surprisingly progressive in a lot of other ways – its plebian "strikes", its multiculturalism, its religious syncretism, its loose sexual morals.

If the resemblance between Roman and modern welfare systems is a mere coincidence, then we have to add a striking number of other coincidences to the list. Eventually the conjunction of all these coincidences starts to look unlikely.

But there is a neat explanation for all of them. States that are militarily secure, economically advanced, multicultural, and urbanized tend to adopt progressive policies (here I am confusingly lumping some values like multiculturalism in as policies, but you know what I mean). Ancient Rome and modern America are both militarily secure, economically advanced, multicultural, and urbanized. In between stand a bunch of countries the Reactionaries like to talk about like the Holy Roman Empire, which were not

militarily secure, economically primitive, monocultural, and more rural. Those countries didn't have progressive policies or values.

The original question was whether ancient Rome could be called a progressive society. I say it was. Scharlach objects that it wasn't, because it didn't have the particular brand of progressive philosophy we do today. But I respond that the philosophy is irrelevant to what we presumably care about – social policies and social outcomes. Policies (like welfare) and outcomes (like the existence of a large class of welfare-dependent poor) were the same in classical Rome and modern America, and for the same reasons. Therefore, it is correct and useful to call classical Rome an early progressive society, though with the obvious caveat that it did not go as far in that direction as our own.

Proposed Biological Explanations for Historical Trends in Crime

My debate on crime rates with Michael Anissimov has been long and meandering, but I think we're starting to come to something of a consensus. I think (I don't know if Michael agrees) that the evidence showing long-term decline in crime from the Middle Ages to the Industrial Revolution is pretty good. There's also irrefutable evidence showing decline in crime from about 1985 to the present. That leaves a gap from about 1850 to 1980.

I previously asserted crime was stable during that period, pointing out similar murder rates between 1850 New York and London and 1980 New York and London, which I trusted more than (say) burglary rates. But Michael replied with a 2002 study showing that improved medical technology has saved a lot of murder victims and bumped their attackers' crimes down to attempted murder, meaning the apparent murder rate is artificially low. Correct that, and murder could have increased by 5-10x or more from 1850 to 1980, which would not be too different from the rates in lesser crimes like burglary.

I am still not entirely certain about this. We have good records on attempted murders for the past 30 years or so, and they have been going down along with the murder rate. And it is surprising that the improvement in medical technology so perfectly balances out the increase in violence. But it's a strong study, and so I will provisionally accept that crime including murder could have risen by 5-10x or more from 1850 to 1980.

But we don't have to accept that the reason is too much democracy or some sort of wacky political point like that.

I have previously come out as a biodeterminist. I suspect most social influences matter less than anyone thinks and most biological influences matter more than anyone thinks. When I say that, everyone always assumes I'm talking about genes, which is too bad because genes are almost the *least* interesting aspect of biodeterminism.

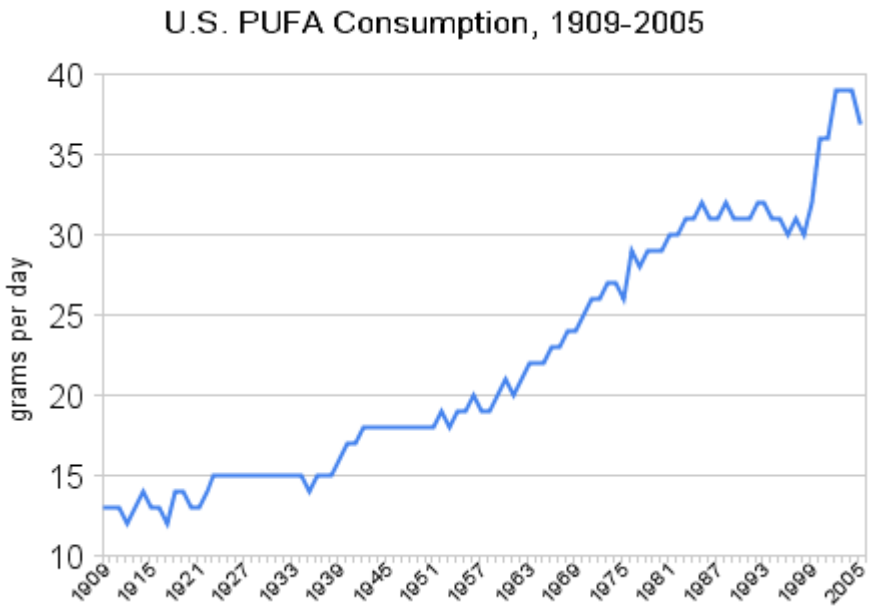
Anyone reading this blog probably already knows that lead is very strongly suspected of causing crime. A generation after gasoline was leaded, crime increased by a factor of four; a generation after lead was banned from gasoline, crime decreased by a factor of four. Levels of automobile lead emissions were found to explain 90% of the variability in violent crime in America. States that banned lead more quickly saw crime drop more quickly. Neighborhoods with higher lead levels consistently had higher crime rates. Blood lead levels show a marked inverse correlation with IQ, and a marked direct correlation with criminal history, even when plausible confounders are taken into account. And neuroscientists have known for decades that lead damages parts of the brain normally involved in good decision-making and in impulse control.

Lead levels started rising with the Industrial Revolution and, although in decline, are still far higher than in pre-industrial societies. They are highest in cities and especially in the inner city. They have shown correlation with crime, teenage pregnancy, and many mental disorders.

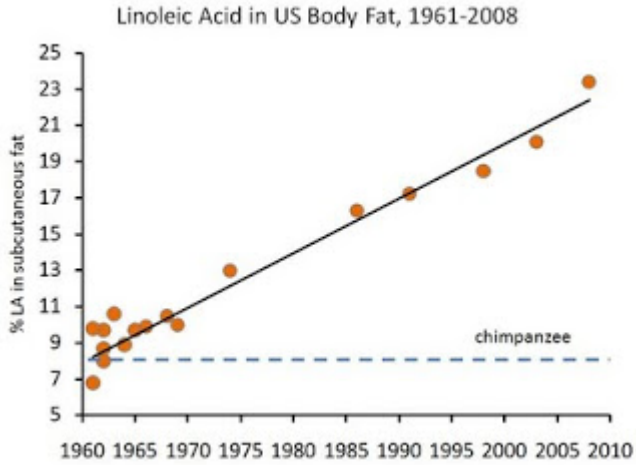
But like I said, everyone reading this blog probably already knows that. So let me talk about something I just learned last week.

Omega-6 fatty acids.

These are some of those "polyunsaturated fatty acid" things you always hear nutrition geeks talking about. They were pretty rare in human diets until the advent of industrial food processing. Here is a mysterious graph for which I have no source:



Here's another that comes from Stephan Guyenet:



So suffice it to say that our consumption of these fatty acids has increased *a lot*. This is not surprising – they are most common in things like the vegetable oil that a bunch of preserved foods have.

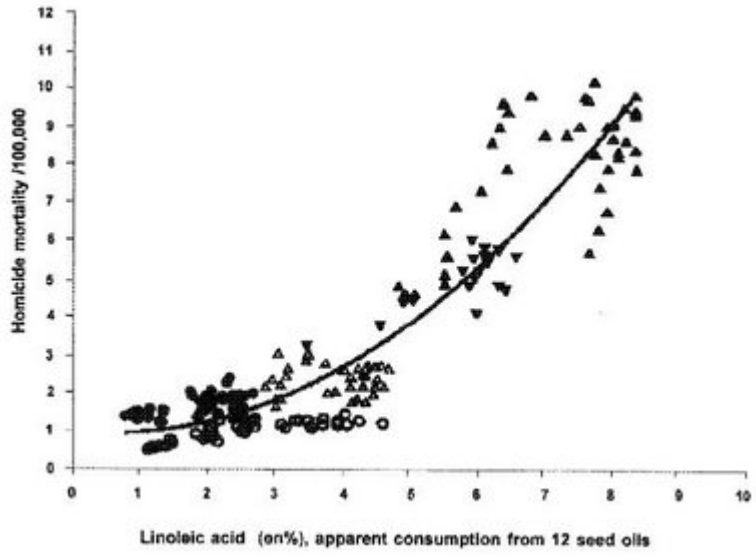
The other main kind of polyunsaturated fatty acid, omega-3, is mostly found in seafood and is the main component of the infamous “fish oil”. It hasn’t increased very much at all and so most people have an abnormally high omega-6:omega-3 ratio compared to the past and to the environment of evolutionary adaptedness.

Omega-3 and omega-6 fatty acids are important for cell membrane fluidity, especially in the brain where they affect neurotransmitter receptors and other neural functions. If there are the wrong amounts of them, this would very plausibly derange various cognitive functions.

So let’s look at Joseph Hibbeln’s paper Seafood Consumption and Homicide Mortality.

The Guardian describes it like so: “Hibbeln and his colleagues have mapped the growth in consumption of omega-6 fatty acids from seed oils in 38 countries since the 1960s against the rise in murder rates over the same period. In all cases there is an unnerving match. As omega-6 goes up, so do homicides in a linear progression. Industrial societies where omega-3 consumption has remained high and omega-6 low because people eat fish, such as Japan, have low rates of murder and depression.”

From Stephen Guyenet’s excellent post Vegetable Oil and Homicide:



I know, I know, it's a nice pretty line, but where are the randomized controlled trials?

To which one answers: "in dozens of different countries around the world". One of the most famous is Gesch et al 2002, which gave dietary supplements including fish oil or placebo to 231 prisoners and found a 25% drop in prison violence (p = 0.03) using intention to treat and 35% (p = .001) using completers. A replication study on 231 Dutch prisoners found almost exactly the same results. Another study of 468 schoolchildren also showed exactly the same results. And… actually, I'm just going to quote from Anatomy of Violence: The Biological Roots of Crime, a book I just found on Google and have suddenly conceived a burning desire to own:

> In Australia, six weeks of omega-3 supplementaion reduced externalizing behavior problems in juveniles with bipolar disorder. In Italy, normal adults taking omega-3 for five weeks showed a significant reduction in aggression compared to controls. In Japan, a randomized controlled trial found that ADHD children with oppositional definat disorder showed a 36% reduction in their oppositional behavior after fifteen weeks of omega-

3. In Thailand, a randomized double-blind trial of the omega-3 fatty acid DHA resulted in a significant reduction in aggression in adult university workers. In the United States, women with borderline personality disorder randomized into supplementation of the fatty acid EPA for two months showed a significant reduction in aggression. Another American study, this time a four-month randomized double-blind placebo-controlled trial of fatty acid supplementation in fifty children, showed a significant 42.7% reduction in conduct-disorder problems.

We have been burned by omega-3 before. Every couple of weeks someone makes an exciting claim about it, and a few weeks later it is shown to be false or overblown. A big government review of the research on mental health basically dismisses everything done thus far as insufficient to draw meaningful conclusions. But I am hopeful.

I will add one more chemical, one of my favorites. Lithium. Many studies (1, 2, 3 find strong (that last one is p = .00003) links between lithium levels in the water supply and an endpoint crime or suicide. Lithium is a known neuroprotective agent, is probably at least calming, and may be otherwise good for the brain.

I am not certain of this, but I have heard from a few sources that modern water treatment/purification removes most minerals, which would suggest we are getting much less lithium than people in the old days who got their water from a well or whatever.

So we are likely getting more lead, more omega-6 (and relatively less omega-3), and less lithium than people in 1850. If there has been an increase in crime and other

undesirable/impulsive behaviors, I think these biological insults are at least as worthy of examination as political changes that have occurred during that time.

Society is Fixed, Biology is Mutable

Today during an otherwise terrible lecture on ADHD I realized something important we get sort of backwards.

There's this stereotype that the Left believes that human characteristics are socially determined, and therefore mutable. And social problems are easy to fix, through things like education and social services and public awareness campaigns and "calling people out", and so we have a responsiblity to fix them, thus radically improving society and making life better for everyone.

But the Right (by now I guess the far right) believes human characteristics are *biologically* determined, and biology is fixed. Therefore we shouldn't bother trying to improve things, and any attempt is just utopianism or "immanentizing the eschaton" or a shady justification for tyranny and busybodyness.

And I think I reject this whole premise.

See, my terrible lecture on ADHD suggested several reasons for the increasing prevalence of the disease. Of these I remember two: the spiritual desert of modern adolescence, and insufficient iron in the diet. And I remember thinking "Man, I hope it's the iron one, because that seems a *lot* easier to fix."

Society is *really hard to change*. We figured drug use was "just" a social problem, and it's *obvious* how to solve social problems, so we gave kids nice little lessons in school about how you should Just Say No. There were advertisements in sports and video games about how Winners Don't Do Drugs. And just in case that didn't work, the cherry on the social engineering sundae was putting all the drug users in jail, where

they would have a lot of time to think about what they'd done and be so moved by the prospect of further punishment that they would come clean.

And that is why, even to this day, nobody uses drugs.

On the other hand, biology is gratifyingly easy to change. Sometimes it's just giving people more iron supplements. But the best example is lead. Banning lead was probably kind of controversial at the time, but in the end some refineries probably had to change their refining process and some gas stations had to put up "UNLEADED" signs and then we were done. And crime dropped like fifty percent in a couple of decades – including many forms of drug abuse.

Saying "Tendency toward drug abuse is primarily determined by fixed brain structure" sounds callous, like you're abandoning drug abusers to die. But maybe it means you can fight the problem head-on instead of forcing kids to attend more and more useless classes where cartoon animals sing about how happy they are not using cocaine.

What about obesity? We put a *lot* of social effort into fighting obesity: labeling foods, banning soda machines from school, banning large sodas from New York, programs in schools to promote healthy eating, doctors chewing people out when they gain weight, the profusion of gyms and Weight Watchers programs, and let's not forget a level of stigma against obese people so strong that I am *constantly* having to deal with their weight-related suicide attempts. As a result, everyone…keeps gaining weight at exactly the same rate they have been for the past couple decades. Wouldn't it be nice if increasing obesity was driven at least in part by changes in the intestinal microbiota that we could reverse through careful antibiotic use? Or by trans-fats?

What about poor school performance? From the social angle, we try No Child Left Behind, Common Core Curriculum, stronger teachers' unions, weaker teachers' unions, more pay for teachers, less pay for teachers, more prayer in school, banning prayer in school, condemning racism, condemning racism even more, et cetera. But the poorest fifth or so of kids show spectacular cognitive gains from multivitamin supplementation, and doctors continue to tell everyone schools should start later so children can get enough sleep and continue to be totally ignored despite strong evidence in favor.

Even the most politically radioactive biological explanation – genetics – doesn't seem that scary to me. The more things turn out to be genetic, the more I support universal funding for implantable contraception that allow people to choose when they do or don't want children – thus breaking the cycle where people too impulsive or confused to use contraception have more children and increase frequency of those undesirable genes. I think I'd have a heck of a lot easier a time changing gene frequency in the population than you would changing people's locus of control or self-efficacy or whatever, even if I wasn't allowed to do anything immoral (except by very silly religious standards of "immoral").

I'm not saying that all problems are purely biological and none are social. But I do worry there's a consensus that biological things are unfixable but social things are easy – or that social solutions are morally unambiguous but biological solutions necessarily monstrous – and so for any given biological/social breakdown of a problem, we figure we might as well put all our resources into attacking the more tractable social side and dismiss the biological side. I think there's a sense in which that's backwards, and in which it's possible to marry scientific rigor with human compassion for the evils of the world.

XI. Social Justice

Practically-a-Book Review: Dying to be Free

I am the last person with a right to complain about Internet articles being too long. But if I did have that right, I think I would exercise it on Dying To Be Free, the Huffington Post's 20,000-word article on the current state of heroin addiction treatment. I feel like it could have been about a quarter the size without losing much.

It's too bad that most people will probably shy away from reading it, because it gets a lot of stuff *really* right.

The article's thesis is also its subtitle: "There's a treatment for heroin addiction that actually works; why aren't we using it?" To save you the obligatory introductory human interest story: that treatment is suboxone. Its active ingredient is the drug buprenorphine, which is kind of like a safer version of methadone. Suboxone is slow-acting, gentle, doesn't really get people high, and is pretty safe as long as you don't go mixing it with weird stuff. People on suboxone don't experience opiate withdrawal and have greatly decreased cravings for heroin. I work at a hospital that's an area leader in suboxone prescription, I've gotten to see it in action, and it's literally a life-saver.

Conventional heroin treatment is abysmal. Rehab centers aren't licensed or regulated and most have little interest in being evidence-based. Many are associated with churches or weird quasi-religious groups like Alcoholics Anonymous. They don't necessarily have doctors or psychologists, and some actively mistrust them. All of this I knew. What I didn't know until reading the article was that – well, it's not just that some of them try to brainwash addicts. It's more that some of

them try to cargo cult brainwashing, do the sorts of things that sound like brainwashing to *them*, without really knowing how brainwashing works assuming it's even a coherent goal to aspire to. Their concept of brainwashing is mostly just creating a really unpleasant environment, yelling at people a lot, enforcing intentionally over-strict rules, and in some cases even having struggle-session-type-things where everyone in the group sits in a circle, scream at the other patients, and tell them they're terrible and disgusting. There's a strong culture of accusing anyone who questions or balks at any of it of just being an addict, or "not really wanting to quit".

I have no problem with "tough love" when it works, but in this case it doesn't. Rehab programs make every effort to obfuscate their effectiveness statistics – I blogged about this before in Part II here – but the best guesses by outside observers is that for a lot of them about 80% to 90% of their graduates relapse within a couple of years. Even this paints too rosy a picture, because it excludes the people who gave up halfway through.

Suboxone treatment isn't perfect, and relapse is still a big problem, but it's a heck of a lot better than most rehabs. Suboxone gives people their dose of opiate and mostly removes the biological half of addiction. There's still the psychological half of addiction – whatever it was that made people want to get high in the first place – but people have a much easier time dealing with that after the biological imperative to get a new dose is gone. Almost all clinical trials have found treatment with methadone or suboxone to be more effective than traditional rehab. Even Cochrane Review, which is notorious for never giving a straight answer to anything besides "more evidence is needed", agrees that methadone and suboxone are effective treatments.

Some people stay on suboxone forever and do just fine – it has few side effects and doesn't interfere with functioning. Other people stay on it until they reach a point in their lives when they feel ready to come off, then taper down slowly under medical supervision, often with good success. It's a good medication, and the growing suspicion it might help treat depression is just icing on the cake.

There are two big roadblocks to wider use of suboxone, and both are enraging.

The first roadblock is the #@$%ing government. They are worried that suboxone, being an opiate, might be addictive, and so doctors might turn into drug pushers. So suboxone is possibly the most highly regulated drug in the United States. If I want to give out OxyContin like candy, I have no limits but the number of pages on my prescription pad. If I want to prescribe you Walter-White-level quantities of methamphetamine for weight loss, nothing is stopping me but common sense. But if I want to give even a single suboxone prescription to a single patient, I have to take a special course on suboxone prescribing, and even then I am limited to only being able to give it to thirty patients a year (eventually rising to one hundred patients when I get more experience with it). The (generally safe) treatment for addiction is more highly regulated than the (very dangerous) addictive drugs it is supposed to replace. Only 3% of doctors bother to jump through all the regulatory hoops, and their hundred-patient limits get saturated almost immediately. As per the laws of suppy and demand, this makes suboxone prescriptions very expensive, and guess what social class most heroin addicts come from? Also, heroin addicts often don't have access to good transportation, which means that if the nearest suboxone provider is thirty miles from their house they're out of luck.

The List Of Reasons To End The Patient Limits On Buprenorphine expands upon and clarifies some of these points.

(in case you think maybe the government just honestly believes the drug is dangerous – nope. You're allowed to prescribe without restriction for any reason except opiate addiction)

The second roadblock is the @#$%ing rehab industry. They hear that suboxone is an opiate, and their religious or quasi-religious fanaticism goes into high gear. "What these people need is Jesus and/or their Nondenominational Higher Power, not more drugs! You're just pushing a new addiction on them! Once an addict, always an addict until they complete their spiritual struggle and come clean!" And so a lot of programs bar suboxone users from participating.

This doesn't sound so bad given the quality of a lot of the programs. Problem is, a lot of these are closely integrated with the social services and legal system. So suppose somebody's doing well on suboxone treatment, and gets in trouble for a drug offense. Could be that they relapsed on heroin one time, could be that they're using something entirely different like cocaine. Judge says go to a treatment program or go to jail. Treatment program says they can't use suboxone. So maybe they go in to deal with their cocaine problem, and by the time they come out they have a cocaine problem *and* a heroin problem.

And…okay, time for a personal story. One of my patients is a homeless man who used to have a heroin problem. He was put on suboxone and it went pretty well. He came back with an alcohol problem, and we wanted to deal with that and his homelessness at the same time. There are these organizations

called three-quarters houses – think “halfway houses” after inflation – that take people with drug problems and give them an insurance-sponsored place to live. But the catch is you can’t be using drugs. And they consider suboxone to be a drug. So of about half a dozen three-quarters houses in the local area, none of them would accept this guy. I called up the one he wanted to go to, said that he really needed a place to stay, said that without this care he was in danger of relapsing into his alcoholism, begged them to accept. They said no drugs. I said I was a doctor, and he had my permission to be on suboxone. They said no drugs. I said that seriously, they were telling me that my DRUG ADDICTED patient who was ADDICTED TO DRUGS couldn’t go to their DRUG ADDICTION center because he was on a medication for treating DRUG ADDICTION? They said that was correct. I hung up in disgust.

So I agree with the pessimistic picture painted by the article. I think we’re ignoring our best treatment option for heroin addiction and I don’t see much sign that this is going to change in the future.

But the health care system not being very good at using medications effectively isn’t news. I also thought this article was interesting because it touches on some of the issues we discuss here a lot:

The value of ritual and community. A lot of the most intelligent conservatives I know base their conservativism on the idea that we can only get good outcomes in “tight communities” that are allowed to violate modern liberal social atomization to build stronger bonds. The Army, which essentially hazes people with boot camp, ritualizes every aspect of their life, then demands strict obedience and ideological conformity, is a good example. I do sometimes

have a lot of respect for this position. But modern rehab programs seem like a really damning counterexample. If you read the article, you will see that this rehabs are trying their best to create a tightly-integrated religiously-inspired community of exactly that sort, and they have abilities to control their members and force their conformity – sometimes in ways that approach outright abuse – that most institutions can't even dream of. But their effectiveness is abysmal. The entire thing is for nothing. I'm not sure whether this represents a basic failure in the idea of tight communities, or whether it just means that you can't force them to exist *ex nihilo* over a couple of months. But I find it interesting.

My love-hate relationship with libertarianism. Also about the rehabs. They're minimally regulated. There's no credentialing process or anything. There are many different kinds, each privately led, and low entry costs to creating a new one. They can be very profitable – pretty much any rehab will cost thousands of dollars, and the big-name ones cost much more. This should be a perfect setup for a hundred different models blooming, experimenting, and then selecting for excellence as consumers drift towards the most effective centers. Instead, we get rampant abuse, charlatanry, and uselessness.

On the other hand, when the government rode in on a white horse to try to fix things, all they did was take the one effective treatment, regulate it practically out of existence, then ride right back out again. So I would be ashamed to be taking either the market's or the state's side here. At this point I think our best option is to ask the paraconsistent logic people to figure out something that's neither government nor not-government, then put that in charge of everything.

Society is fixed, biology is mutable. People have tried *everything* to fix drug abuse. Being harsh and sending drug users to jail. Being nice and sending them to nice treatment centers that focus on rehabilitation. Old timey religion where fire-and-brimstone preachers talk about how Jesus wants them to stay off drugs. Flaky New Age religion where counselors tell you about how drug abuse is keeping you from your true self. Government programs. University programs. Private programs. Giving people money. Fining people money. Being unusually nice. Being unusually mean. More social support. Less social support. This school of therapy. That school of therapy. What works is just giving people a chemical to saturate the brain receptor directly. We know it works. The studies show it works. And we're still collectively beating our heads against the wall of finding a social solution.

Drug Testing Welfare Users is a Sham, But Not for the Reasons You Think

Some people say the War on Drugs is 'unwinnable'. But there's actually a foolproof solution that cures drug addiction approximately 100% of the time. That solution is – put people on welfare in Tennessee.

Or at least that is what I am led to believe by articles like Mic's A Shocking Thing Happened When Tennesee Decided To Drug Test Its Welfare Recipients, which describes said shocking thing as:

> 1 out of 812 applicants tested positive for drugs. One. Single. Person. Tennessee conservatives suspicious that welfare recipients are a bunch of drug-addicted slackers were proven dead wrong. Big surprise!
>
> After instituting dehumanizing drug-testing requirements to welfare recipients on July 1, 10 people total were flagged for possible drug use and asked to submit to testing. Five others tested negative, and four were rejected after refusing. As Think Progress notes, that means that just 0.12% of all people applying for cash assistance in Tennessee have tested positive for drugs, compared to the 8% who have reported using drugs in the past month among the state's general population. If you assume the four people who refused were on drugs, it's still a paltry 0.61%.
>
> In other words, the plan intended to verify right-wing beliefs that welfare recipients are a bunch of drug-addicted slackers looking for a handout has demonstrated exactly the opposite.

The article has 11,000 notes on Tumblr right now, I've seen it all over my Facebook feed as well, and the same story has been taken up, with the same editorial line, by a host of other news sources. Jezebel: State Drug Program Busts A Whopping 37 Welfare Applicants. Wall Street Journal: Few Welfare Applicants Caught In

Drug Screening Net So Far. New Republic: Red States' New Tax On The Poor. Daily Kos: Tennessee Just Wasted A Lot Of Money Drug Testing Welfare Recipients. ReverbPress: Another GOP Fail: 0.2% Of Tennessee Welfare Recipients Found To Use Illegal Drugs. Mommyish: Results Of State Drug Testing Prove Gross Assumptions About Welfare Applicants Are Wrong. Washington Post: Scott Walker's Yellow Politics.

These stories all make the point that we have many stereotypes about the poor, and one such stereotype is that the use lots of drugs, but in fact these sorts of welfare programs find them to use fewer drugs than the general population, and therefore we should stop being so prejudiced.

And if they were found to use only two-thirds, or half as many drugs as the general population, this might indeed be the lesson.

But look at the numbers in the quoted Mic article. Welfare users use only about *one percent* as many drugs as the general population. *Really?*

No. Not really at all. According to legitimate research in this area, poor people use as many drugs as anyone else and probably more. The National Household Survey on Drug Abuse found that illegal drug use was slightly higher in families on government assistance (9.6%) than families not on government assistance (6.8%). The National Coalition For The Homeless notes that about 26% of them use drugs, which is about 2.5x as high as the general population. I crunched some data I have from the hospital I work at, and it shows that poor people (defined as people who get health insurance through an aid program) have moderately higher rates of drug use related problems than the general population. So these articles are reporting a drug use rate in the Tennessee population about one percent of that ever reported in any comparable poor population anywhere else.

Kate from Gruntled and Hinged brings up another curious inconsistency. The false positive rate for drug tests is – well, it depends on the test procedure, but it's usually at least 1%. So if

every single welfare user in Tennessee was 100% clean, we would *still* expect between 1% to 5% positive drug tests. Instead, they got 0.12% positive drug tests. This isn't just suspiciously good, it's *impossibly* good.

So what's going on here?

Before I explain, here's a collage of the stock photos displayed above some of those news stories I linked to.

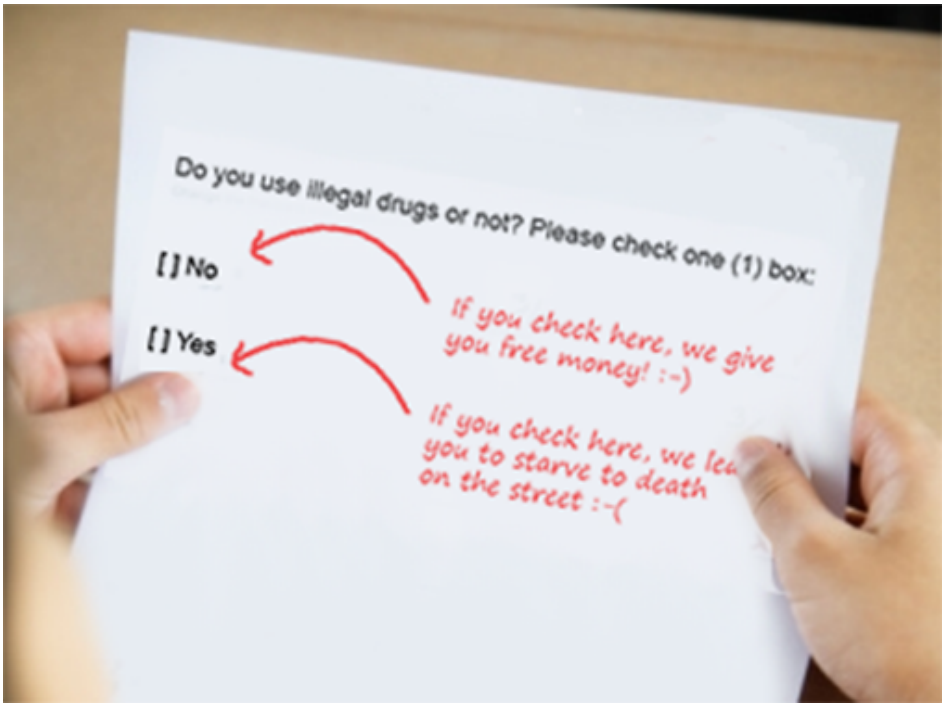
I now have a picture on my website called urine_collage.png

If you're familiar with the state of the American media, you won't be surprised to learn that urine was not involved in the ovewhelming majority of this program's drug tests.

So how did they test people for drugs?

They gave them a written test, where the test question was basically "do you use illegal drugs or not?" You can see the exact procedure on the sidebar here.

And lo and behold, the overwhelming majority of people answered that they didn't.



A more accurate stock photo they could have used

Now the numbers make sense. It's not that only 0.2% of welfare recipients use drugs. All this tells us, if anything, is that 0.2% of welfare recipients are on so many drugs they can't figure out how to check "NO" on a form.

Why would the government do something like this? As best I can tell, the plan was originally to give everyone urine checks, but in Florida the courts decided that urine-checking people without prior suspicion was unconstitutional. The Republicans were pretty attached to their "drug test welfare recipients" plan and didn't want to look like they were wimps who backed down just because of one little court case, so they decided to give people the written test in the hopes of having prior suspicion for the people who said yes. Sure, it made no sense, but they could still tell their constituents they were drug testing those welfare recipients, and *in principle* they'd won an important victory. Or something.

Which raises another interesting question – how did Florida's urine-based program do before the courts struck it down?

According to the media, abysmally. MSNBC: Drug Testing Welfare Recipients Looks Even Worse, "[Florida Governor] Scott's policy

was an embarrassing flop. Only about 2 percent of applicants tested positive, and Florida actually lost money". TBO: Welfare Drug Testing Yields 2% Positive Results, "Newton said that's proof the drug-testing program is based on a stereotype, not hard facts." ATTN: Why Drug Testing Poor People Is A Waste Of Time And Money, "Florida tested welfare recipients for four months before its drug test mandate was thrown out by the courts. Only 2.6 percent of welfare recipients tested positive. The rest of the Florida's population use drugs at a rate of 8 percent. So, again, welfare recipients used drugs less than everyone else."

Now we're merely at one-quarter of the drug use rate people with good methodologies find. Improvement!

So I looked up exactly how this works. Apparently welfare recipients were asked to pay for their own drug tests, and would be reimbursed if the results came back negative. 7000 welfare users did this, but 1600 declined to do so – numbers that were not mentioned in most of the pieces above.

Opponents of the program say that maybe those 1600 people could not find drug testing centers near them, or couldn't afford to pay for the tests even with the promise of reimbursement later, or something like that. I am sure that some of them did indeed decline for reasons like those.

But also, people on welfare don't have very much money [citation needed]. If I were a welfare recipient, and they were going to drug test me and not reimburse me if I came out positive, and I was on drugs, I would decline the hell out of that test.

Suppose that the poor in Florida use drugs at the same rate as the poor in various studies and surveys – about 10%. We have 8600 welfare recipients, so we would expect 860 drug users. Of the 7000 who agreed to testing, we know that 2.5% are drug users – that's 175 people. That in turn would suggest that of the 1600 who refused testing, about 685 were drug users – 40% or so. That would imply that about 80% of drug users versus about 12% of nonusers refused testing.

These numbers seem pretty reasonable to me. Most welfare users want to keep their benefits, so the majority will agree to testing, but a few will inevitably fall through the cracks because they can't reach a testing center or because they have moral objections to the tests. On the other hand, clued-in drug users will realize that for them, testing means a major inconvenience and monetary charge without any likely corresponding gain. So we would expect drug users to decline testing at a higher rate than nonusers. In order to use the Florida data to say that welfare recipients *in general* use drugs at a rate of 2%, we would need to assume that drug users were no more likely to refuse drug testing than nonusers, even though the testing rewarded non-use with money but punished use with a loss of money.

(note that there are some different numbers in different places for Florida. I assume that these represent different years, stages of testing, parts of Florida, etc, but I'm not sure. The only one that is *seriously* different from what I'm saying above is the one that says "only 1% of people declined testing". After some search, I'm pretty sure that's referring to that only 1% of people made appointments for testing, then cancelled later. But I am less confident in the Florida numbers than in the analysis of Tennessee)

So the Florida numbers are consistent with welfare recipients using drugs less, more, or the same amount as the general population.

So I have a question for you guys.

How come Brian Williams is being dragged over the coals for lying in the media, but everyone who publishes these kinds of articles gets off scot-free?

If I understand correctly, Williams said that his helicopter got shot at when he was in Iraq, but in reality he was just in a helicopter in Iraq at the same time as some other helicopter nearby was getting shot at. This is obviously stretching the truth, but it seems to me it could have been worse. No important policy decisions are going to hinge upon exactly which helicopter Brian Williams was in. And he didn't

get it *infinitely* wrong – for example, there was, in some sense, a war in Iraq.

On the other hand, discussions of how many poor people use drugs is pretty important for all sorts of policy questions, and these people *completely* dropped the ball. So why does nobody get reprimanded for this kind of thing?

You might argue that Brian Williams' actions were obviously malicious and deceitful, but that screwing up drug numbers is an excusable mistake. I say it's exactly the opposite. Brian Williams did exactly what I unfortunately do all the time – unthinkingly tell a story the much cooler way it should have happened, the way it happened in my head – rather than the way it actually did happen (my colleagues elsewhere in the psychiatry blogosphere go further and call this "normal brain function").

On the other hand, I have more trouble imagining a situation in which I would accept the claim "only 0.1% of poor people use drugs, which is barely one percent of the rate in the general population" without wanting to do a little more research to see if it is true. If your reporters are capable of making this mistake honestly, *get better reporters*.

But I'm not sure it's honest. A lot of these sources admit they took their story from a Think Progress piece on the issue. Think Progress *does* mention that the tests are a sham, although only in one sentence that is easy to miss. Either the secondary reporters didn't read Think Progress thoroughly, or they consciously decided not to mention it.

But even if it was an honest mistake, I still have trouble excusing their arrogance. I mean look at that Jezebel article. The writer says this proves that people who think welfare recipients use drugs "consider 'facts' troublesome" and that their "entire social philosophy boils down to 'Ew, poor people.'"

You're saying that's not as bad as a helicopter-related embellishment?

Yes, okay, drug testing welfare applicants is in fact probably a bad idea. It's a bad idea because the courts have banned doing it in a way more effective than asking them politely if they use drugs or not, but it was a bad idea even before that. It's a bad idea because drug tests have frequent false positives, but it's a bad idea even without that. It's a bad idea because quitting drugs is really hard and denying people benefits isn't going to help.

But if, in the service of proving this to be a bad idea, you decide it's acceptable to fudge the numbers to make your point, horrible things happen. First, you contribute to a culture of telling lies and lose the opportunity to protest when the other side does it. Second, you make it harder to trust you on anything else.

But most important, tell one lie and the truth is forever after your enemy. I recently argued that we need to reform suboxone prescribing laws, because it's the best anti-addiction medicine we've got and right now poor people can't access it. . Why should anyone listen to me now? They can just answer "Actually, that would be a waste of money. As per an article I read in Jezebel, pretty much no poor person has ever been addicted to drugs." Then the laws don't get reformed and people die.

The Meditation on Creepiness

As far as I know there aren't a lot of areas where feminists and pickup artists are natural allies, but I can think of one person they would both despise equally. And he has a special place in my heart.

I can't quite remember his name and Google doesn't help, but let's call him al-Fulani. al-Fulani was a classical Islamic poet. When he was a young man traveling the world, he stopped by an oasis town to gather water for his camel and there he passed by a young woman. They exchanged a Significant Look, but said nothing to one another, and in the morning he left the oasis and never saw her again. But he was so impressed by her beauty that he spent the rest of his life composing poems to and about her, which according to the story I heard became among the most exquisite works of Arabic literature even though Google turns up exactly zero of them and maybe I dreamt this entire thing.

The pickup artists would call this "one-itis" and say he had no "game" since he was obsessing over this one woman instead of "playing the field". The feminists would say he was a "rape-y creep". And actually, they're both right. al-Fulani's behavior was neither a healthy way to satisfy his own needs nor fair to the poor woman he fixated on. *Rationally* it's stupid and horrible. Rationally Dante was stupid and horrible for fixating on Beatrice, Romeo was stupid and horrible for fixating on Juliet, and pretty much every love affair in literature up until the 20th century when people switched to writing books where antiheros slept with a bunch of women but never felt anything for any of them until finally they

Developed Ennui - *rationally* all those love affairs were stupid and horrible. They assume that romantic attraction by some crazy form of magic.

But sometimes the magic works. The first time future President Lyndon Johnson met Lady Bird he asked her out on a fancy date; she was shocked at the presumptousness but accepted, later saying she felt "drawn to him like a moth to a flame". On that first date, less than twenty-four hours after they met, he proposed marriage to her. When she said 'of course not are you crazy' he started calling her and writing letters to her practically nonstop; ten weeks later she finally agreed. LBJ tried to insist the wedding occur that same day; Lady Bird managed to bargain him down to "tomorrow". They were married the next day and then had a perfect idyllic relationship that lasted the next forty years until LBJ's death.

I am friends with several married people like LBJ. Sometimes both spouses just knew from the moment they saw each other that it was meant to be. Sometimes only one of them did, and certain amounts of pestering and wooing and opinion-changing were necessary. Sometimes those certain amounts were very high. Most of these couples tend to be older people. A few are my age but conservative Christians. A few are neither old nor old-fashioned but just awesome people.

I am also friends with Normal Proper People. If LBJ or his female equivalent tried to propose to them on the first date, they'd scream at him to get the hell away from them, then post about it on a "What Was Your Worst First Date Ever?" thread on Reddit. Then they'd go to a party, get drunk, make out with someone on the couch, realize a few weeks later that they were kind of sort of dating them and might as well continue, and

after two to four years of "going steady" they'd get married because that's what you do after dating someone for two to four years. A few years later, they would have an affair with their personal trainer who was younger and better-looking. Plus or minus a marriage and personal trainer affair, these seem to be the majority of the people my age whom I know.

And what got me thinking about this was a comment on that Less Wrong thread that got me thinking about this whole gender thing to begin with. I want to make it clear I am not mocking or criticizing this comment and that it is a perfectly rational way to behave and actually much more rational than the way I am behaving. It says:

> *Actually, I have run into enough guys who treat me like I'm the last woman on earth because I'm a female nerd that I've developed an aversion to anything resembling that type of behavior. I was understanding about their enthusiasm at first, because I want a nerd, too, but it just doesn't work to date someone when they're acting like you're their last chance. They want to move too fast, they create expectations, they become biased and won't hear me when I talk about things that may be incompatibilities. That intensity throws a wrench into the process of getting to know someone. I grok their sense of necessity about being careful in how they present themselves, and I approve of this thread (There are a lot of things I wish I could say to guys - we need to communicate, and I have been wishing for an opportunity to do that), but on the individual level, I am easily spooked by signs of early attachment, overly optimistic probability estimates about us working out, and impatience to see signs of an established connection. I go on the alert for these signs of*

irrationality if a person treats me "like a celebrity" or similar.

I am pretty sure I have never met this particular woman, but I have certainly been the kind of guy she is talking about. I used to operate through Burning Life-Consuming Crushes, usually initiated in the first few days I met someone, and if I'd had LBJ's courage and awesomeness I would have asked any one of them to marry me and totally gone through with it if they said yes. Oddly enough (or not, if you've read Malcolm Gladwell's *Blink* or the more reputable studies in the same genres) these first impressions were almost always correct, I found these people to be physically and mentally and emotionally compatible with me, I became good friends with most of them, and quite honestly I would probably still marry some of them after a few minutes' thought if they asked me tomorrow.

Eventually I was socialized into the Correct Way To Feel Attraction, which is "Huh, I guess this girl is pretty cute. I'll invite her out, and if she says no, then no big deal because that girl there is pretty cute too." This is what happened with my first girlfriend. She was a wonderful woman and I have nothing whatsoever bad to say about her, but I asked her out kind of knowing that the relationship would be enjoyable and then fizzle out, and sure enough the relationship was enjoyable and then fizzled out. This was probably exactly why she was my first girlfriend: it gave me the non-desperate-looking-ness that helped me seem attractive to her[1].

So this seems to be another Rule of Intergender Communication like the two I mentioned in the last post: “Don’t come on too strong”.

But if women make a policy of excluding guys who show strong feelings for them, then logically they will end up with either guys who have only a vague and temporary preference for them, or Machiavellian liars.

I’ve tried the Machiavellian liar routine a few times myself. “Oh, hey, you’re Jennifer or Jessica or Julia or whatever, right? I appear to have totally by coincidence ended up at this table with you. Anyway, you seem kind of okay. Want to go out to dinner sometime? Saturday’s no good because I have things to do that night.” Meanwhile in my head I’m going over what we’re going to name our children.

It’s pretty hard to maintain and it’s also really unpleasant and it also makes me feel like a horrible person and it also means that if I ever do get into a relationship with Jennifer or Jessica it will be based on deception and lies and probably continue that way (“It’s our six month anniversary! Can I get her the beautiful personalized gift that will make her super-happy and so make me super-happy as a result, or would that be creepy and I should just get her some crappy half-dead flowers instead?”). Even if I pull it off, I will be doing an imperfect simulation of what a guy who really doesn’t care much for her could do perfectly, and so I will be strictly inferior to him.

Probably most men know they can't manage it, don't even try, and end up independently re-inventing the courtly love tradition: admiring an unattainable woman from afar and showering her with presents as an expression of their transcendent yet hopeless love. Or, as we moderns call it, being a Nice Guy (TM) and therefore Worse Than Hitler (TM).

So I think these filters work and people who have a policy of rejecting suitors who really deeply desire them in a way that makes them not interchangeable with the next "prospect" to come along - they will, in fact, successfully eliminate suitors who really deeply desire them and consider them non-interchangeable. And then ten years later one night in bed they ask their personal trainer why their husband or wife is so frigid.

I know that the Official Narrative is that you're supposed to not get too obsessed with someone until you've been in a relationship with them a while, and you ask them out when you just have a vague preference for them but later you warm up to them and after a few months or years you're genuinely in love and *then* you can do all the stuff I want to do immediately like write them sonnets and sestinas and maybe some ruba'iyat.

But the Official Narrative doesn't take into account that *actually* when I like someone my brain tells me right away and goes into Full Obsession Mode. Maybe there are people who don't work like that. Maybe they're the ones who write

Official Narratives, while the rest of us are wasting our time writing sestinas and exquisite works of Arabic literature.

Now, don't get me wrong. I know that True Love is really inconvenient. It might not be requited, and then it would be a huge mess and no one would have any idea what to do, because our culture tells us that True Love Must Always Conquer Everything. If some woman I didn't like expressed True Love for me, it would make me feel guilty and horrible.

And because I'm just as susceptible to the Just World Fallacy as anyone else, I would tell them it wasn't true love at all but just plain Creepiness. And that it makes her a bad person and she should be ashamed of herself and so rejecting her is not only okay but *actively heroic*. And all my neighbors would support me in this, because we all know that True Love is the most powerful thing in the universe, even more powerful than nuclear weapons, and so we can't just let random people go around *having it* any more than we would just let random people have the Bomb[2].

But when we reach the point where letting it slip that you love someone is pretty much social suicide, that's…not good. I'm trying to imagine what G. K. Chesterton would write if he saw that sentence above - "I know that True Love is really inconvenient" - and then write *that*, but I'm no G. K. Chesterton and also everything Chesterton wrote was beautiful but totally illogical and I don't want to end up like that anyway.

It may be I'm itching to channel Chesterton because I *am* saying something illogical. If I had to support all this with an argument developed by my rational side rather than my Islamic-poetry-reading side, it would look something like this:

1. A sudden intuitive obsession with another person as a romantic partner ("True Love") is often accurate, as shown both by data (eg the sort of stuff you see in *Blink*) and by anecdote (eg LBJ).

2. It is also really really awkward when it happens so[3] mainstream modern culture has developed a norm of keeping it inside and punishing people who express it. Most people will specifically avoid anyone who tries to show True Love.

3. Unfortunately, this selects against people who have strong romantic preferences, who are probably also the people who are most likely to make good relationship partners.

4. People are afraid of a social norm that they *have to* accept anyone who declares True Love for them, and obviously that would be a bad social norm. Declaring True Love should not force the object of affection to reciprocate and maybe should not even count in the person's favor.

5. But it shouldn't count *against* the person either, and you shouldn't actively *penalize* the person for looking like they Truly Love you.

6. If you do, you may well end up with a partner who doesn't Truly Love you. Maybe they will come to love you anyway as your relationship blossoms, but it seems less certain they if they did at the start.

But I'm pretty sure that's all motivated thinking. It's definitely not my True Objection. My True Objection is an aesthetic appreciation for the fiery dazzling love that comes out of nowhere. It's a sense of crushing ugliness when I consider the modern culture of "Let's meet for coffee sometime, or not, meh, plenty of fish in the sea, so whatever." It's one of those base-level preferences that can't be CEVed away any more than romance itself could. If you don't share the preference that's fine, but I wish you wouldn't make life so difficult for people who do.

1: *Actually, I should expand upon that word "desperation". I've been told it's really non-sexy, because it implies you need this girl to say yes because you're not cool enough to get any other. But another possible explanation is that you don't *want* another and that not all human beings are interchangeable to you. And this really ought to be a point in your favor.*

2: *Well, sort of. It seems to me that there is a certain kind of self-consciously suave and obviously false True Love which is socially acceptable, typified in a singer crooning "You're the only one for me, baby." I can't put my finger on the difference between that and the al-Fulani type of True Love, but I'm pretty sure it's there and detectable by a third party.*

3: *I expect there's probably also a signaling explanation for why True Love isn't tolerated. Maybe if anyone were allowed to show True Love, everyone would fake it and there would be an arms race or something? I can't put my finger on it right now, but I bet it's a good one. On the other hand, I'm not sure it's good enough. Banning the expression of True Love seems supervillainish enough that it's hard to imagine what could justify it.*

Actually, I think I support a more general Supervillain Test: if a supervillain were plotting a specific social change, would we assemble a band of scrappy yet loveable teenagers with mysterious powers to thwart him? If yes, we should want to thwart the change even if it happens organically as a result of impersonal forces.

The Meditation on Superweapons

Let's talk about the US missile defense shield.

Right now it can only shoot down a few missiles some of the time. But maybe one day it will be able to shoot down many missiles all of the time. The balance of power between the United States and Russia depends on mutually assured destruction. For either country to gain the ability to shoot down many missiles all of the time would upset this balance. Therefore, Russia opposes the US missile defense shield.

The United States tries to reassure Russia. "We're just building this shield to protect ourselves from Iran and North Korea", they say. This is super reasonable. The United States really does face a serious threat from Iran and North Korea. Building a missile defense shield is a great idea for reasons that have nothing to do with Russia. If Russia starts threatening to attack the United States if they don't stop building their shield, Russia looks like an aggressive jerk meddling in matters that don't concern it.

But say the United States finishes its defense shield, and then happens to disagree with Russia over some minor issue like the Syria conflict. "I think you better do what we say," says America. "We could crush you like a bug." And Russia says "But you told us your shield had nothing to do with us!". And the US answers "And we were telling the truth. We didn't intend it against you. But here we are, disagreeing with you and having a spare superweapon. It wasn't our original intent. But now, we own you."

Now let's talk about anti-Semitism.

Suppose you were a Jew in old-timey Eastern Europe. The big news story is about a Jewish man who killed a Christian child. As far as you can tell the story is true. It's just disappointing that everyone who tells it is describing it as "A Jew killed a Christian kid today". You don't want to make a big deal over this, because no one is saying anything objectionable like "And so all Jews are evil". Besides you'd hate to inject identity politics into this obvious tragedy. It just sort of makes you uncomfortable.

The next day you hear that the local priest is giving a sermon on how the Jews killed Christ. This statement seems historically plausible, and it's part of the Christian religion, and no one is implying it says anything about the Jews today. You'd hate to be the guy who barges in and tries to tell the Christians what Biblical facts they can and can't include in their sermons just because they offend you. It would make you an annoying busybody. So again you just get uncomfortable.

The next day you hear people complain about the greedy Jewish bankers who are ruining the world economy. And really a disproportionate number of bankers are Jewish, and bankers really do seem to be the source of a lot of economic problems. It seems kind of pedantic to interrupt every conversation with "But also some bankers are Christian, or Muslim, and even though a disproportionate number of bankers are Jewish that doesn't mean the Jewish bankers are disproportionately active in ruining the world economy compared to their numbers." So again you stay uncomfortable.

Then the next day you hear people complain about Israeli

atrocities in Palestine, which is of course terribly anachronistic if you're in old-timey Eastern Europe but let's roll with it. You understand that the Israelis really do commit some terrible acts. On the other hand, when people start talking about "Jewish atrocities" and "the need to protect Gentiles from Jewish rapacity" and "laws to stop all this horrible stuff the Jews are doing", you just feel worried, even though you personally are not doing any horrible stuff and maybe they even have good reasons for phrasing it that way.

Then the next day you get in a business dispute with your neighbor. If it's typical of the sort of thing that happened in this era, you loaned him some money and he doesn't feel like paying you back. He tells you you'd better just give up, admit he is in the right, and apologize to him - because if the conflict escalated everyone would take his side because he is a Christian and you are a Jew. And everyone knows that Jews victimize Christians and are basically child-murdering Christ-killing economy-ruining atrocity-committing scum.

He has a point - not about the scum, but about that everyone would take his side. Like the Russians in the missile defense example above, you have allowed your opponents to build a superweapon. Only this time it is a conceptual superweapon rather than a physical one. The superweapon is the memeplex in which Jews are always in the wrong. It's a set of pattern-matching templates, cliches, and applause lights.

The Eastern European Christians did not necessarily have evil intent in creating their superweapon, any more than the Americans had evil intent in their missile shield. No particular action of theirs was objectionable - they were genuinely worried about that one murder, they were genuinely worried

about Israeli atrocities. But like the Americans, once they have that superweapon they can use it on anyone and so even if you are a good person you are screwed.

This rule of "never let anyone build a conceptual superweapon that might get used against you" seems to be the impetus behind a lot of social justice movements. For example, it's eye-rollingly annoying whenever the Council on American - Islamic Relations condemns a news report on the latest terrorist atrocity for making too big a deal that the terrorists were Islamic (what? this bombing just killed however many people, and all you can think of to get upset about is that the newspaper mentioned the guy screamed 'Allahu akbar' first?), but I interpret their actions as trying to prevent the construction of a conceptual superweapon against Islam (or possibly to dismantle one that already exists). Like the Jew whose best option would have been to attack potentially anti-Jewish statements *even when they were reasonable in context*, CAIR can't just trust that no one will use the anti-Muslim sentiment against non-threatening Muslims. As long as there are stupid little trivial disputes between Muslims and non-Muslims over anything at all, that giant anti-Muslim superweapon sitting in the corner is just too tempting to refuse.

This is also one reason (of at least three) why I have serious reservations about feminism.

Sometimes I read feminist blogs. A common experience is that by the end of the article I am enraged and want to make a snarky comment, so I re-read the essay to pick out the juiciest quotes to tear apart. I re-read it and I re-read it again and eventually I find that everything it says is both factually true and morally unobjectionable. They very rarely say anything

silly like “And therefore all men, even the ones who aren’t actively committing this offense I’m arguing against, are evil”, and it’s usually not even particularly implied. I feel like the Jew in the story above, who admits that it’s really bad the Jewish guy killed the Christian child, and would hate to say, like a jerk, that Christians aren’t allowed to talk about it.

But like him I am uncomfortable. Like him I can’t shake the worry that they are building a conceptual superweapon that could be used against me.

Feminism is a memeplex that provides a bunch of pattern-matching opportunities where a man is in the wrong and a woman is in the right. To give a very personal example, I mentioned a few days ago how I was close friends with a woman until I asked her out and she then decided to have a fit and cut off all contact with me. Normally everyone would agree I was in the right and try to console me and maybe even her own friends would tell her she was overreacting. But thanks to feminism she has a superweapon - she can accuse me of being a Nice Guy (TM) and therefore Worse Than Hitler (TM). The appropriate cliche having been conveniently provided, enough people decide to round to the nearest cliche and decide that I am in the wrong that the incident raises her status and decreases my own.

And aside from my own experience I just keep seeing the superweapon turned on innocents. The awkward guy who asks a woman out in the wrong way, who to me is a figure of pity, gets superweaponed and turned into a figure of public vituperation. When a woman gives a guy a bunch of obvious hints and so he tries to kiss her or something and then gets yelled at and called a creep, he can’t protest “I’m really sorry,

but she was giving me a bunch of obvious hints" or else he will get superweaponed and everyone will pattern-match him to a rapist. And if anyone disagrees with the feminist position on any political issue, from free contraception to affirmative action, then even if they have reasonable philosophical arguments they get superweaponed and everyone completes the pattern as "misogynist".

Or in general, everyone agrees we need to do a certain number of things to deal with prejudice against women, but people generally disagree on *exactly* how far we should go, and if any two people disagree the one who supports less action risks superweaponing.

Also, whenever someone accuses feminists of being trigger-happy with their superweapon, they tend to turn their superweapon on the accuser. It creates kind of a vicious cycle.

Now the feminists would say that I too have a superweapon called "patriarchy", and that they're just continuing the arms race. This is true, but it doesn't lead to a stable state like what the guns rights advocates claim would happen if everyone had guns where we would all be super-polite because nobody wants to offend a guy who's probably packing heat. It leads to something more like a postapocalyptic anarchy where everyone has guns and we're all shooting each other. If there's a conflict between a man and a woman, and the people involved happen to be old-fashioned patriarchalist types, then the man will automatically win and everyone will hate the woman for being a slut or a bitch or whatever. If there's a conflict between a man and a woman, and the people involved happen to be feminists who are familiar with the memeplex and all its pattern-matching suggests, then the woman will

probably win and everyone will hate the man for being a creep or a bigot or whatever. At no point does everyone become respectful and say “Hey, we’re all reasonable people with superweapons, let’s judge this case on its merits instead of pattern-matching to the closest atrocity committed by someone of the same gender”.

It also seems to me that the patriarchy is sort of an accident, where men ruled because they were big and strong and couldn’t imagine doing otherwise and their values just sort of coalesced over time, and the struggle seems to be getting them to realize it’s there. Whereas the feminists know all about discourse and power relations and so on and are quite gung ho about it and they’re staying up late at night reading books with titles like *How To Build A Much Deadlier Superweapon* (I assume this book exists and is written by Nikola Tesla).

I’m all for mutual superweapon disarmament, but I’m not sure I like the whole mutually assured destruction thing as much. My history, and I think the history of a lot of people who are liberal and pro-choice and so on and so forth but really wary of feminism and social justice - is that we spent our college years totally supporting social justice and helping out in the superweapon factories because it’s our duty to fight rape and racism and so on and since *we* were nice respectful people obviously the superweapon would never be used on *us*. Then we got in some kind of trivial disagreement with a woman or a minority or someone, or we didn’t want to go far enough. Then they turned the superweapon on us, and it was kind of a moment of “wait, this was sort of the plan all along, wasn’t it?”

I have an ambiguous respect for the white males who continue

to be serious parts of the feminist and anti-racist movement - not just “well obviously I’m against discrimination but I’m not sure I’m drinking your Kool-Aid” people like myself, but the sort of who major in the appropriate college courses and write for the blogs and totally identify with the movement. It’s ambiguous because I’m not sure if they’re really naive (“Oh, they would *never* use this superweapon unless they had a *really really good reason*, and certainly not on the *nice people* like me”) or whether they are really selfless (“I know this superweapon will eventually be used against me and other innocent people, but it’s so important to arm this group against real enemies that I will help them build it anyway for purely consequentialist reasons.”)

But I myself am not going there. The United States has mostly reassured Russia by promising them that their missile shield will be able to deflect the few and weak Korean/Iranian missiles it might face, but not the more numerous and more advanced Russian varieties. I think it’s probably possible to create forms of social justice that would actually be focused against real threats and not provide free superweapons to anyone who wants to vaporize a few unattractive nerds before dinner. I just don’t think the community in its current form is very good at pursuing them.

The Meditation on the War on Applause Lights

Suppose the President gets asked to veto a new hydroelectric dam. After thinking it over, he does so. He says the dam would destroy the environment and flood many homes. And the people ask him "Why do you hate Italians so much?"

What?

Well, if there's no hydroelectric plant, they have to make that energy some other way. A lot of it will be from fossil fuels. Fossil fuels contribute to global warming. Global warming raises sea levels. Rising sea levels are destroying Venice. The destruction of Venice will end the livelihoods of thousands of Italians. So if the President vetoes the dam, the best explanation is that he hates Italian people.

This isn't just an example of not using the Principle of Charity. No one uses the Principle of Charity. I push the Principle of Charity endlessly and think it's the greatest thing since sliced bread and toy with tattooing "Principle of Charity, people!" in big letters on my chest so that whenever people go on one of their demonization trips I can just take off my shirt and they'll be like "Oh, sorry". And even *I* forget to use the Principle of Charity most of the time, because it's really hard. But this is something way more malignant. This is like an active Principle of Anti-Charity here.

The Principle of Anti-Charity can do anything. No matter what the President's next move is, we can make that part of the War On Italians as well. Does the President cut the military budget?

The US is the core of NATO; any decrease in US forces will have to be compensated by the other NATO countries if the alliance is to stay strong. Therefore Italy has to invest more money in defense, dealing another blow to its already crumbling economy. Shame on the President and his Italian-hating ways.

Or maybe the President raises the military budget. This would probably mean an expansion of existing military bases. Some of those are in Italy, and every time they expand the Italians nearby protest what the New York Times describes as "traffic congestion, environmental damage, and the possibility of terrorist attacks." So the President clearly wants terrorists to attack Italy.

So maybe the President just refuses to even touch the military budget at all. Well, in that case he's weak and passive and a bad leader, and probably no one will ever build a monument to him. And most monuments are built out of marble. And the best marble comes from Cararra in Italy. He must trying to sabotage the Italian marble industry!

(I mean, when the President makes one anti-Italian decision, you can kind of put your head in the sand and believe it might be a coincidence. But when *all* of his decisions hurt Italy in some way? Hardly!)

So the Principle of Anti-Charity is pretty hard to disprove. The reason I get so exasperated when anyone talks about gender is that the Principle of Anti-Charity seems to be the Official Standard For Debate. Here I will be quoting from The Uncredible Hallq, which is actually a really awesome blog with great analysis of some issues in philosophy of religion;

despite me having a problem with this one minor thing I absolutely recommend it:

> *When you look at stuff like the reaction to Todd Aikin saying rape victims don't need abortions because they won't get pregnant if it's a "legitimate rape," what you see is people waking up to the fact that the anti-abortion movement isn't about their public rhetoric about "partial birth abortion." It's full of vile extremists who want to deny women their basic right to bodily autonomy.*

I find this fascinating. Here is this one guy[1] whom 99.999% of people *on the anti-abortion side* have condemned and tried to distance themselves from. Every single prominent Republican from Mitt Romney to Sean Hannity to John Ashcroft condemned him, which is almost unpredecented in terms of Republicans condemning fellow Republicans. The head of the RNC decided to ban him from the Republican Convention and called him "stupid". Republican Super PACs and the party itself stopped funding his race. Karl Rove publicly threatened to murder him - he sounded like he meant it figuratively, but since it's Karl Rove he should probably keep his doors bolted just in case.

A few people have said something like how they think he is a great guy personally and share his views about abortion *even though that particular comment were idiotic*, and a few people say they think the media reaction was disproportionate *even though his comments were idiotic*. One or two really fringe extremist pro-life groups have said they kind of agree with

him *although his way of putting it was idiotic*. This is as close as anyone came to saying he wasn't an idiot.

And so of course we naturally conclude from this that he has spoken the secret heart of the anti-abortion movement and his opinions can be considered representative.

But more importantly, let's go to the last sentence of the quote: "It's full of vile extremists who want to deny women their basic right to bodily autonomy."

They "want" to deny women their basic right to bodily autonomy in the same way that the President "wants" to destroy the livelihood of Italians. That is, they support a policy for completely different reasons and it will end out denying women their rights. If you would feel awkward saying the President is plotting to drown the Venetians, please feel exactly as awkward saying conservatives are plotting to deny women their right to bodily autonomy.

The same is true of the contraceptive mandate. Its Wikipedia article includes quotes like:

> *Sen. Frank Lautenberg (D-NJ) said Republicans "want to take us back to the Dark Ages ... when women were property."*
>
> *The Democratic Congressional Campaign Committee says, "House Republicans have launched an all-out war*

on women since taking the Speaker's gavel over a year ago."

Right. It's obviously all about women. Because Republicans are usually so thrilled about government forcing them to do things against their religion. And they just *love* when Obama pushes through health care mandates.

The War On Women is exactly as real as the War On Christmas. People do not launch Wars On Applause Lights. People sometimes *accuse their opponents* of launching Wars on Applause Lights, because then instead of having to argue that a new hydroelectric dam is really necessary, they can just sit and watch while the President has to defends himself against hordes of angry Italian-American voters.

Speaking of Wars on Things, let's talk about the War on Terror. Everyone agrees terrorism is really bad. Some people want a Strong Response To Terror, which in practice consists of waterboarding some people and then bombing a randomly chosen Middle Eastern country. Other people want a More Measured Response To Terror, which in practice consists of trying to figure out what kind of things we do that make us a target for terrorism and then not doing them.

The former group of people call the latter group of people Soft On Terror. I think it's a terrible phrase, but it could be worse. They could accuse them of being part of "terrorism culture", an all-pervading belief system that terrorism is somewhere between excusable and admirable, and that every time they

vote against a new drone bombing it is because they secretly think terrorists are great people. And every time they try to figure out the conditions that promote terrorism and decrease them, it's because of their deep-seated desire to blame the victims of terrorism for the attacks.

The point of this post is not for me to say either side is correct, or even that one side is not completely barking up the wrong tree and their so-called "solutions" are actually exactly the wrong way to go about it and will just make the terrorists stronger. Please do not try to infer my position in the actual debate just because I am talking about the meta-debate. If you have to know, I agree with the moderate liberal position on terrorism and I agree with the feminist position on the issue that this is an obvious metaphor for. But that doesn't matter.

What matters is that this is also a Principle of Anti-Charity issue. If you hear that some Democratic Senator voted not to invade Iraq, and your first thought is "I bet he secretly loves terrorists and thinks the victims of terrorism deserve what they get", then your head is not screwed on straight.

Suppose you hear Noam Chomsky say that maybe one way to decrease terrorist attacks would be to stop propping up the Saudi royal family. And maybe you know he's wrong and you have study after study showing that terrorists don't care about the Saudi royal family and that actually countries that support the Saudi royal family less are even more likely to be attacked by terrorists. But nevertheless if you decide that it's *totally impossible* that he's just a nice person who honestly wants to

help - if you decide the *only* explanation for Chomsky's actions is that he's Osama bin Laden's best pal and secretly goes out to the cemetery every night to dance on the grave of 9-11 victims - if you use his advice as proof that our society is really a pro-al-Qaeda "terrorism culture" - then you have left the Way.

People do not launch Wars On Applause Lights. People do not Secretly Love Boo Lights. If you keep it up like this maybe I am seriously going to have to get that Principle of Charity tattoo.

FOOTNOTE: What the heck was Akin thinking, anyway? To anyone familiar with cognitive biases, the answer should be obvious. It's the just-world fallacy and the eternally springing hope that policy debates should be one-sided. Suppose you believed abortion was genuinely murder and just as bad as killing a grown adult. In that case, if women could get pregnant from rape, you would have to make an impossible moral choice between committing murder and forcing a woman to bear her rapist's child. It would be horrible and you would feel like a monster whichever you did. And the world is just and fair and never presents you with horrible impossible moral choices, therefore women cannot get pregnant from rape. So when he read a (terrible, unethical) doctor who claimed exactly that in a (terrible, unethical) article published in a real (terrible, unethical) book, he gave a big sigh of relief and didn't think twice about it.

The Meditation on Superweapons and Bingo

I usually blog about a mix of philosophy, medicine, and random things that go on in my personal life. According to my LiveJournal Statistics page, a typical blog post of mine from last month when I was blogging every day and about writing really interesting stuff like meeting a guy possessed by demons got eight hundred page views per day by about a hundred fifty LiveJournal users a day. As soon as I started writing about gender, it shot up to about twenty-five hundred page views by three hundred fifty users a day. On the one hand, I like popularity as much as anyone else. On the other hand, I feel like by writing on a hot-button issue and taking a side on the object-level debate, I'm kind of doing something sort of dirty; like now I'm only one or two levels above those blogs that write "The Democrats suck, because they love Big Government! LOL!" and get a million subscribers a day. So I will make one final object-level post today, a meta-level post tomorrow, and then try to limit myself to at absolute most one culture war per week from now on.

I

Sometimes people complain that it's scary how oblivious the other side is to their arguments. But I know something scarier.

On r/atheism, a Christian-turned-atheist once described an "apologetics" group at his old church. The pastor would bring in a simplified straw-man version of a common atheist argument, they'd take turns mocking it ("Oh my god, he said that monkeys can give birth to humans! That's hilarious!") and then they'd all have a good laugh together. Later, when they met an actual atheist who was trying to explain evolution to them, they wouldn't sit and evaluate it dispassionately. They'd pattern-match back to the ridiculous argument they heard at church, and instead of listening they'd be thinking "Hahaha, atheists really *are* that hilariously stupid!"

Of course, it's not only Christians who do that. I hear atheists

repeat the old “I believe the Bible because God said it was true. We know He said it was true because it’s in the Bible. And I believe the Bible because God said it is true” line *constantly* and grin as if they’ve said something knee-slappingly funny. I’ve never in my entire life heard a Christian use this reasoning. I have heard Christians use the “truth-telling thing” argument sometimes (we should believe the Bible because the Bible is correct about many things that can be proven independently, this vouches for the veracity of the whole book, and therefore we should believe it even when it can’t be independently proven) many times. If you’re familiar enough with the atheist version, and uncharitable enough to Christians, you will pattern-match, miss the subtle difference, and be thinking “Hahaha, Christians really *are* as hilariously stupid as all my atheist friends say!”

Sometimes even the straw-man argument is unnecessary. All you need to do is get in a group and make the other side’s argument a figure of fun.

There are lots of good arguments against libertarianism. I have collected some of them into a very long document which remains the most popular thing I’ve ever written. But when I hear liberals discuss libertarianism, they very often head in the same direction. They make a silly face and say “Durned guv’mint needs to stay off my land!” And then all the other liberals who are with them laugh uproariously. And then when a real libertarian shows up and makes a real libertarian argument, a liberal will adopt his posture, try to mimic his tone of voice, and say “Durned guv’ment needs to stay off my land! Hahaha!” And all the other liberals will think “Hahaha, libertarians really *are* that stupid!”

Many of you will recognize this as much like the Myers Shuffle. As long as a bunch of atheists get together and laugh at religious people who ask them to read theology before criticizing it, and as long as they have an easily recognizable name for the object of their hilarity like "Courtier's Reply", then whenever a religious person asks them to familiarize themselves with theology the atheist can just say "Courtier's Reply!" and all the other atheists will crack up and think "Hahaha, religious people really *are* that stupid!" and they gain status and the theist loses status and at no point do they have to even consider responding to the theist's objection.

This tendency reaches its most florid manifestation in the "ideological bingo games". See for example "Skeptical Sexist Bingo", feminist bingo, libertarian troll bingo, anti-Zionist bingo, pro-Zionist bingo, and so on. If you Google for these you can find thousands, which is too bad because *every single person who makes one of these is going to Hell*.

Let's look at the fourth one, "Anti-Zionist Bingo." Say that you mention something bad Israel is doing, someone else accuses you of being anti-Semitic, and you correct them that no, not all criticism of Israel is necessarily anti-Semitic and you're worried about the increasing tendency to spin it that way.

And they say "Hahahahahhaa he totally did it, he used the 'all criticism of Israel gets labeled anti-Semitic' argument, people totally use that as a real argument hahahaha they really are that stupid, I get 'B1' on my stupid stereotypical critics of Israel bingo!"

You say "Uh, look, I'm not really sure what you're getting at. I

recognize that there is real anti-Semitism and I am just as opposed to it as you are but surely when when see the state excusing acts of violence against Palestinians in the West Bank we…"

And they say "Hahahhaha G1, I got G1, he pulled the old 'I abhor real anti-Semitism' line this is great, guys come over here and look at what this guy is doing he's just totally parroting all the old arguments every anti-Semite uses!"

So it may be scary when your opponent is unaware of your arguments, but it is much scarier when your opponent has a sort of vague dreamlike awareness of your arguments, which immediately pattern-match cached thoughts about how horrible a person you would have to be to make them.

But this is still not the scariest thing.

Because if your opponent brings out the Bingo card, you can just tell them exactly what I am saying here. You can explain to the pro-Israel person that they are pattern-matching your responses, that you don't know what strawman anti-Zionist they're thinking of but that you have legitimate reasons for believing what you do and you request a fair hearing, and that if they do not repent of their knee-slapping pattern-matching Bingo-making ways *they are going to Hell*.

No, the scariest thing would be if one of those bingo cards had, in the free space in the middle: "You are just pattern-matching my responses. I swear that I have something legitimate to tell you which is not just a rehash of the straw-man arguments you've heard before, so please just keep an open mind and hear me out."

If someone did *that*, even Origen would have to admit they were beyond any hope of salvation. Any conceivable attempt to explain their error would be met with a "Hahahaha he did the 'stop-pattern matching I'm not a strawman I'm not an inhuman monster STOP FILLING OUT YOUR DAMN BINGO CARD' thing again! He's so hilarious, just like all those other 'stop-pattern matching I am not a strawman' people whom we know only say that because they are inhuman monsters!"

But surely no one could be that far gone, right?

Listen:

"I'm not racist, but…"

If you are like everyone else on the Internet, your immediate response is "Whoever is saying that is obviously a racisty racist who loves racism! I can't believe he *literally* used the 'I'm not racist, but…' line in those exact words! The old INRB! I've got to get home as fast as I can to write about this on my blog and tell everyone I really *met* one of those people!"

But why would someone use INRB? It sounds to me like what they are saying is: "Look. I know what I am saying is going to sound racist to you. You're going to jump to the conclusion that I'm a racist and not hear me out. In fact, maybe you've been trained to assume that the only reason anyone could possibly assert it is racism and to pattern-match this position to a racist straw man version. But I actually have a non-racist reason for saying it. Please please *please* for the love of Truth

and Beauty just this one time throw away your prejudgments and your Bingo card and just listen to what I'm going to say with an open mind."

And so you reply "Hahahaha! He really used the 'look I know what I'm saying is going to sound racist to you you're going to jump to the conclusion that I'm a racist and not hear me out in fact maybe you've been trained to assume…' line! What a racist! Point and laugh, everyone! POINT AND LAUGH!"

And of course "sexist" works just as well as "racist" here, even though the latter is more familiar.

This is what I mean by "conceptual superweapon". *This* is what it looks like to stare into the barrel of a gigantic lunar-based death ray and abandon all hope. This is why I find feminism and the social justice community in general so scary.

II

Let's switch topics. Let's switch to medical testing. Although Medical Testing For Biochemists is complicated and involves scary words like "pharmacokinetics", Medical Testing for Doctors is much easier and goes like this:

A Magic Mystery Box fell to Earth during an eerie thunderstorm. If we wave the Magic Mystery Box over a patient, it beeps and displays a number from one to one hundred. Now sometimes low-numbered patients have cancer, and sometimes high-numbered patients are healthy, but in general the higher the number the more likely the patient is to have cancer. Sometimes.

Suppose the doctor has two choices. She can refer the patient to surgery, where surgeons will cut him open, look to see if there really is a cancer, and if so try to take it out. This surgery is expensive, unpleasant, and there's always the chance the surgeon's hand will slip and cut something important and the patient will die.

Or the doctor can say "Oh, you don't really have cancer" and do nothing.

If she tells a patient who has cancer that he's healthy, the patient will die, sue the doctor, or both. If she tells a patient who is healthy that he needs to go to surgery for further cancer investigation, she makes the patient needlessly terrified, wastes the surgeon's time, risks complications from the surgery, and costs the health system thousands of dollars.

So she waves the Magic Mystery Box over the patient, and it beeps and says "22". Now what?

In practice, doctors establish a threshold. The threshold will be a number like "40". If the test is above 40, the patient gets surgery. If the test is below 40, they send the patient home.

How does one choose the right threshold? A low threshold means means that doctors will catch almost all cancer, but they'll also end up sending a lot of healthy people for dangerous unnecessary surgery. A high threshold means that few healthy patients will ever suffer the risks of unnecessary surgery, but probably a lot of cancer will go undetected.

If the surgery is really dangerous but the cancer isn't that bad, the doctors will choose a high threshold. If the surgery is quick

and safe but the uncaught cancer would be fatal, the doctors will choose a low threshold. But no matter what number they choose, all they can do is minimize the harm. Unless they sent every single patient of theirs to surgery, there will always be a few cancers that are uncaught. Unless they never send anyone to surgery, there will always be a few false alarms. As long as the test itself is imperfect, the doctors' decision will always unfairly harm a few patients. They just need to figure the threshold that harms as few as possible.

(If you're familiar with statistics, you already recognize this situation as Type I and Type II errors. If you're familiar with utilitarianism, you already recognize the solution as setting the threshold to maximize total utility across all patients. I'm not saying anything new here.)

If a doctor uses the established thresholds and refers a patient to surgery that turns out to be unnecessary, there are laws preventing that patient from suing her. The same is true if all the tests came back below the threshold, she said he was fine, and he later turned out to be super unlucky and have a totally undetectable form of cancer. The doctor did everything right. She just got unlucky. Those laws are *really good*. If they didn't exist, it would either be impossible to practice medicine, or else doctors would be optimizing for not being sued rather than for doing good medicine even more than they already are. If they only existed in one direction (eg doctors who did unnecessary surgeries couldn't be sued, but doctors who missed cancer could), that would be even worse - any doctor not heroic enough to go against her own self-interest would refer every patient to surgery.

Politics is nowhere near as rational as medicine. Politicians

don’t think in terms of thresholds. No one ever says “The more regulations we put on businesses, the fewer customers will get scammed by shady con men. But also the more likely it is that we unnecessarily penalize honest businesses. So we need to find the threshold value that minimizes the total unfairness to businesses and customers.” Instead they say either “We need to fight for more regulations and anyone who says otherwise is in the pay of Big Business!” or “We need to cut through all the red tape and anyone who says otherwise is in the pay of Big Government!”

No one ever says “The more restrictions we place on welfare, the more certain we’ll be that no one is abusing the system. But the more restrictions we place on welfare, the more certain we will be that some poor people who desperately need it can’t get it. Therefore, we should determine the relative disutilities of people defrauding us and of needy people not being able to use the system, and act to maximize total utility.” Instead they say “Anyone who opposes tight welfare restrictions is a welfare queen trying to scam you!” or “Anyone who wants any welfare restrictions hates poor people!”

Gender issues also involve thresholds.

A man who wants to know whether it is okay to ask a woman out can try to read her social cues and appeal to lists of known social norms. This is his Magic Mystery Box. Sometimes it works. Sometimes it doesn’t. He needs to set a threshold for action: how open to an advance does she has to look before he asks her or flirts with her or whatever. If the threshold is too low, he will be a creep and she will feel harrassed. If the threshold is too high, no one will ever ask anyone else out and everyone will die alone and unloved.

Another man is in love. He wants to know if he can express his love to a woman without worrying that it is "creepy" or "coming on too strong". Again social cues give him a Magic Mystery Box. Again he must set a threshold. If the threshold is too low, he will end up creeping people out. If the threshold is too high, then no one can ever be in love and all couples must be formed by deciding the other person is good-looking and so you will settle for them.

We need to dismantle social structures that favor men aka patriarchy. It's kind of hard to figure out which ones those are - is the preponderance of male math professors because of the patriarchy, or something else? We can run studies and surveys of women in the math field and try to get some preliminary conclusions - our Magic Mystery Box. But we need some threshold for intervention like fixing a quota of 50% women mathematicians in every college. If our threshold is too low, we end up with tokenism and promoting unqualified people. If our threshold is too high, we end up perpetuating the patriarchy.

Some men (and women!) express political positions whose consequences could hurt women. It's unclear whether they support those positions because they honestly believe they are good for society, or because they are evil people who deliberately aim at misogyny. We can psychoanalyze them - our Magic Mystery Box - but we must decide a threshold. If the threshold is too low, evil misogynists can get away with their evil misogyny and no one will call them on it. If the threshold is too high, we will end up demonizing a bunch of random people, giving feminism a reputation as "those people who go around demonizing innocents", and totally destroy any

chance at friendly political discourse.

No threshold should ever be set at zero. If a doctor sets the Magic Mystery Box threshold to zero, then she will end up referring every single patient for dangerous surgery. “Doctor, I’ve been having a bit of a sore throat these past few….SURGERY! NOW!”

But if one side has a superweapon, it’s impossible to argue for the other. If the threshold starts at forty, and one doctor says “But we can’t be the sorts of monsters who would refuse a potential cancer patient live-saving surgery!”, and this argument is a deeply-ingrained part of medical culture and the other doctors don’t want to be tarred as cancer-sympathizers, then the threshold goes to 30. Then another doctor brings up the same argument, and the threshold goes to 20. Soon the threshold is at zero and they’re referring rashes and hay fever for surgery and no one can protest because they don’t want to look Pro-Cancer.

If it is impossible to ever say “You know, the social justice people make some good points, but on this issue here they’ve gone too far,” then the threshold on all of those questions above just keeps inching downward until it hits zero.

And if people are punished for their results rather than their actions - if you can get called a creep even though you did your best to take her hints and followed all the rules - then that’s like only suing the doctors who miss cancer. It’s going to bring the threshold down to the zero “operate on everyone” level even faster.

III

When I Googled for good examples of those bingo games to post above, it was pretty hard to find the Zionism ones and so on. Almost every ideological bingo game out there was feminist. This is not a coincidence.

For those who have absorbed the associated memes, feminism is a fully general conceptual superweapon. It has attempted and probably completed the task of making every possible counterargument so unthinkable that any feminist can refute it just by reciting the appropriate bingo square, then pointing and laughing.

If a man thinks women are less oppressed than she claims, she can say "male privilege!" and point and laugh.

If a man thinks there are some areas where the threshold has moved too far toward women, she can make a grave expression and intone "What About Teh Menz?" (now the name of a major blog, which is actually pretty good) and point and laugh.

If a man thinks parts of the reason why some men are jerks toward women is because women actually are more likely to date jerks than people who are respectful, she can gleefully declare "You're a Nice Guy (TM) and therefore Worse Than Hitler (TM)!" and point and laugh.

If a man tries to explain his own perspective to her or provide any alternative theory to men-being-horrible, she can say he's "mansplaining again!" and point and laugh.

If a man asks not to be immediately pattern-matched to the

nearest hostile cliche when he tries to present his opinion, she can say he's using a variant of the old "I'm not sexist, but…" line. And point. And laugh.

During the past few days, some people have criticized me for nonstandard use of terms. They have tried to tell me that the legitimate definition of a feminist term isn't a bingo-square demonization that can be used to shut down debate, but [complex legitimate reasonable definition]. Well, okay. I agree all of these words have possible legitimate definitions and uses and were created for good reasons. The same is true of the word "Communist". It means a person who supports a classless society with common ownership of all goods. This has nothing to do with "communist" the way it is used in actual American political discourse eg "Obama is a communist because he wants universal health care!" If I criticize the Republicans for using the word "communism" as a debate-stopper, saying "But in this here dictionary Communist has a legitimate and useful meaning" is not a response. People created the word because it was useful and meaningful. Then it got picked up and placed into the fuel chamber of a superweapon. Now it is ten million degrees and radioactive and bears no resemblance to its former self. You might not find the terms above used in exactly the way above in the Official Oxford Dictionary Of Gender Relations, but I did check Google and urbandictionary.com to make sure that I wasn't completely generalizing from my own experience here.

My view on feminism isn't really driven by my view on gender relations or women or men or society. It's driven by my view on applause lights, on inability to urge restraint, on death spirals, on anti-charity, on zero-threshold medical testing, on superweapons, and most of all on epistemic hygiene. I don't

care how righteous your cause is, you don’t get a superweapon so powerful it can pre-emptively vaporize any possible counterargument including the one asking you to please turn off your superweapon and listen for just a second. No one should be able to do that.

I apologize for this post being so long. I wanted to make sure it wouldn’t fit on a bingo board.

An Analysis of the Formalist Account of Power Relations in Democratic Societies

[**Epistemic Status** | *Sooooorta re-inventing the wheel here. Nevertheless, I feel I deserve tenure at a major university for managing to write an essay with this title. Somebody please make this happen.*]

If Donald Trump and Rebecca Black got in a bar fight, who would win?

(Don't just answer "society". This is a serious question which will illuminate structures of dominance in modern culture.)

In the short-term, Donald Trump would easily beat up Rebecca Black. He's bigger, manlier, and it should be pretty easy for him to overpower a teenage girl.

In the medium-term, the ensuing media circus would be entirely in Rebecca's favor. No matter who started the fight or how justified their *casus belli*, the media would portray it as "Donald Trump beats up a little girl". The media optimizes for outrage, and "arrogant billionaire beats up poor sympathetic teenage girl" is more outrageous than "Poor sympathetic teenage girl rabidly attacks arrogant billionaire". Besides, Trump is a confirmed Person Whom It Is Fun To Dislike, and it seems very unlikely that a media mogul would receive angry self-righteous letters to the editor for picking on him. Rebecca could basically walk into a bar where Donald is drinking quietly, smash a chair over his head for no reason, and the media would *still* find a way to make sure it ended with him coming under irresistable pressure to apologize to her on

national TV.

In the long-term, the media circus would die down. Trump would still live in a gigantic mansion from which he controls large parts of the world economy, and Rebecca Black would still be a B- or C- list celebrity desperately trying to avoid having everyone forget her.

So which of the two of them has more *power*?

If I correctly understand Mencius Moldbug, which is always a big 'if', I think he is arguing that the title goes uncontroversially to Ms. Black. From Unqualified Reservations:

> *"The truth is that the weapons of 'activism' are not weapons which the weak can use against the strong. They are weapons the strong can use against the weak. When the weak try to use them against the strong, the outcome is… well… suicidal.*
>
> *Who was stronger - Dr. King, or Bull Connor? Well, we have a pretty good test for who was stronger. Who won? In the real story, overdogs win. Who had the full force of the world's strongest government on his side? Who had a small-town police force staffed with backward hicks? In the real story, overdogs win.*
>
> *'Civil disobedience' is no more than a way for the overdog to say to the underdog: I am so strong that you cannot enforce your 'laws' upon me. I am strong and might makes right - I give you the law, not you me. Don't think the losing party in this conflict didn't try its own*

‘civil disobedience.’ And even its own ‘active measures.’ Which availed them - what? Quod licet Jovi, non licet bovi.

In the real world in which we live, the weak had better know their own weakness. If they would gather their strength, do it! But without fighting, even ‘civil disobedience.’ To break a law is to fight. Those who fight had better be strong. Those who are not strong, had better not fight.

And this is how Chomskyism killed Aaron Swartz and may yet get its hands on a similar figure, Julian Assange. You know, when I read that Assange had his hands on a huge dump of DoD and State documents, I figured we would never see those cables. Sure enough, the first thing he released was some DoD material.

Why? Well, obviously, Assange knew the score. He knew that Arlington is weak and Georgetown is strong. He knew that he could tweak Arlington’s nose all day long and party on it, making big friends in high society, and no one would even think about reaching out and touching him. Or so I thought.

In fact, my cynicism was unjustified. In fact, Assange turned out to be a true believer, not a canny schemer. He was not content to wield his sword against the usual devils of the Chomsky narrative. Oh no, the poor fscker believed that he was actually there to take on the actual powers that be. Who are actually, of course, unlike the cartoon villains… strong. If he didn’t know that… he knows it now!

> *Better to be a live dog than a dead hero. But had Aaron Swartz plugged his laptop into the Exxon internal network and downloaded everything Beelzebub knows about fracking, he would be a live hero to this day. Why? Because no ambitious Federal prosecutor in the 21st century would see a route to career success through hounding some activist at Exxon's behest. Your prosecutor would have to actually believe he was living in the Chomsky world. Which he can't, because that narrative is completely inconsistent with the real world he goes to work in every day."*

I can think of at least two different problems with this passage.

The first is that it's outright false. Moldbug later uses the example of pro-lifers protesting abortion as an example of an unsympathetic and genuinely powerless cause. Yet as far as I can tell abortion protesters and Exxon Mobile protesters are treated more or less the same. In both cases, polite protesters who stick to the law are allowed to keep doing their thing, or occasionally get arrested and then immediately released, but those who actually hurt people or damage property are punished.

The second is that, even if it were true, it would be taking an overly simplistic view of "real power". Moldbug says we can determine the real power based on who wins. But what kind of winning? There are kinds of winning where you beat someone in a bar fight. There's the kind of winning where you get such

overwhelming support of public opinion you can force them to apologize to you on TV. And there's the kind of winning where you go home to Trump Tower at night.

Suppose Rebecca Black starts a barfight with Donald Trump, the media spins it as sympathetic to Black and excuses her actions, and Trump ends up with egg on his face. Does that make Black more powerful than Trump?

Or to put it another way, suppose I throw my shoe at the President, and everyone is sympathetic to me, and the President suggests not pressing charges in order to look merciful, and the government is under lots of political pressure to pardon me. Does this make me *more powerful* than the President?

Or to put it another way, suppose I am a liberal activist lobbyist who says lots of mean things about ExxonMobil is and is a constant thorn in their side. I spend my entire life harassing them through bringing legal cases against them and convincing Congress to pass laws against them. I win all my legal cases, blocking some of their drilling, and Congress passes all the laws I want, raising their tax rate a little. Whenever ExxonMobil tries to condemn me in any way, there is a huge political outcry and they back off. Does this make me more powerful than ExxonMobil?

No. What I described would be pretty successful for a life of activism. But in the end, ExxonMobil is going to just drill somewhere else, and figure out some tax shelter policy that

completely avoids whatever law I got Congress to pass against them. In the end, they will still be very rich and control the world economy, and I will probably get some award and feel good about myself but make zero difference. In the end, I'm the one winning the media circus, and they're the one going home to Trump Tower.

There are kinds of power where you lose every single fight you get into, maybe on purpose, and still end out more powerful than before, because the direction your power is growing is orthogonal to the direction people are fighting you in, or because the actual power structure is buried much too deeply for the theater of public relations to even notice. Indeed, this is the only kind of power worth having.

We will call this sort of gather-your-power-bit-by-bit-and-hide-it-places-no-one-knows sort of advantage that ExxonMobil and Donald Trump have *structural power*, and the sort of win-at-media-circuses-and-maybe-trials advantage that environmental activists and Rebecca Black have *social power*. An equally good term would be *unconscious power* and *conscious power*, because wherever anyone makes a conscious decision they will happily decide in favor of the environmentalists and Ms. Black, and it is only the unconscious non-decisions that skew the real world in favor of ExxonMobil and Mr. Trump.

Both Moldbug and liberal activists seem to understand this distinction sometimes, although other times they can be bizarrely pigheaded about conflating the two types of power.

Moldbug's shtick as I interpret it claims that social power should be more in line with structural power. Liberal activists seem to think that structural power needs to change and social power can change it.

Taking Silver In The Oppression Olympics

Here is another of my favorite graphs

The solid gray line is white people rating how much discrimination they think there was against black people at different periods. The dashed gray line is white people rating how much discrimination they think there was against white people at different periods. We see that the average white believes that around the year 2000 there started to be more discrimination in America against white people than against black people.

If we extrapolate - which would be kind of irresponsible from this study as it is retrospective, but humor me - it looks like quite soon, and maybe even today since the graph is several years old, that the average white person will actually feel more discriminated against than the average black person does.

The people on the Reddit thread pretty much used this to conclude that white people are dumb and should never be allowed to talk about race.

I think that might be part of it but also that there is a more subtle problem. Social power is much easier to notice than structural power, especially if you're not the one on the wrong end of the structural power.

To give a very timely example, every February there's this boring low-level repetitive argument about "Why is there a Black History Month but not a White History Month?" "No, *every* month is White History Month, that's the whole reason a Black History Month is necessary." Even if the latter statement is true, it's a lot easier to notice that black people get an Officially Endorsed Month (social/conscious power) than that white people tend to come off better during the eleven theoretically neutral months (structural/unconscious power).

Or to give another example, there are Official Laws saying that women should be privileged over men in some sorts of employment and college admission determinations; anyone who claimed that men should be officially privileged over women by law in any field would be ostracized (social/conscious power). On the other hand, actual hiring decisions tend to favor men over women, and this is mediated by subconscious assessments of competence (structural/unconscious power).

As I said before, I bet I'm reinventing the wheel here and somebody else has come up with this idea long ago and given it a different name that I just don't recognize (it seems possible that "privilege" might just be a really horrible failed attempt at raising awareness of unconscious/structural power)

The Obvious Liberal and Conservative Responses

But even if this is well-trodden ground, I have yet to hear anyone on either side give their respective obvious responses.

The Obvious Liberal Response is this: We like claiming that activists and minorities are powerless and oppressed. And we can see why the fact that they really have all the social/conscious power could be jarring, and even upsetting to very literal-type people with unrealistically high expectations for how honest discourse is supposed to be.

But this doesn't make us *wrong*. Social/conscious power, in and of itself, is kind of a booby prize. Having a History Month dedicated to your race is not a terminal goal.

The things people actually care about, like money, success, influence, and psychological health, come entirely from structural/unconscious power. A city may spend your tax money on colorful "We Love Minorities And Want More Of Them" posters, but if the mayor and all five city councillors are straight white men, then not only are the straight white men not oppressed *on net*, but they're not even suffering in any discernible way *at all*.

The *only* point of having social/conscious power is to try to influence the distribution of structural/unconscious power.

Social/conscious power is a lever that can be used to move structural/unconscious power.

So the goal in distributing social/conscious power isn't to give everyone an exactly equal amount, the way a nice but naive person might expect. The goal in distributing social/conscious power is to distribute it in whatever way causes everyone to end up with an equal amount of structural/unconscious power. Since straight white men continue to be winning the structural/unconscious power game, no matter how unfairly biased the social/conscious power is toward genderqueer minority women, it's obviously *not biased enough*.

If someone had told me this was the liberal argument ten years ago, it would have saved me a crazy amount of hand-wringing. But there's a missing conservative argument too, and that would be this:

Okay, we've been trying for let's say fifty years to use social/conscious power as a lever to move structural/unconscious power.

Just to use race as an example, fifty years ago, there were explicit laws keeping black people down, and scientific racists in universities were blithely speculating on the cranial capacity of "Negroids" without a second thought. Today, an impressive amount of the Western world's academic output by weight is now devoted to yelling about how much we hate racism and homophobia. We have successfully reached the point where a single ambiguously racist comment can bring down pretty

much any politician in the country, and where people who use the word “fuck” like it’s going out of style are terrified even to quote, let alone use, ethnic slurs. In terms of progress in deploying social power against racism, we have come pretty darned far.

Yet the black/white income gap, which is probably the best objective measure we have of structural/unconscious power, worse today than forty years ago when good records first started being kept. Fifty years of feminists telling people to rape less has resulted in a trend line for rape that looks exactly like that for every other violent crime. The biggest success of the anti-inequality movement, higher incomes for women, seems to be an economic transition that had only a little to do with any kind of a social justice movement (citation admittedly needed, but that’d be a whole post in itself).

So what if social/conscious power *just isn’t that good a lever?* We know that in at least in a business environment, promoting diversity has zero positive impact and in fact may just make people more racist. If this is true on a social level, it would fit nicely with the stagnant/disimproving structural/unconscious power situation despite the vastly improved social/conscious power situation.

This makes the last sentence of the liberal argument above sound suddenly terrifying. “Since straight white men continue to be winning the structural/unconscious power game, no matter how unfairly biased the social/conscious power is toward genderqueer minority women, it’s obviously *not biased enough*.” Although biasing the social/conscious power

situation toward minority groups is not nearly as big a disaster as my conservative friends seem to think, I don't think it's completely effect-less either, especially if the results from the business case continue to apply and the more people talk about racism the more racist people become.

Combining the conservative contention "Giving more social/conscious power doesn't increase structural/unconscious power" with the liberal contention "We need to keep giving more social/conscious power until the structural/unconscious power increases to the right level" means that we will just end up giving infinite amounts of social/conscious power, to no positive effect. This, the conservative might argue, would at the very least be an inefficient use of resources, not to mention such an easy and attractive solution that it would prevent us from looking for things that do have an effect.

And Back To The Original Question

So I think the Moldbuggian paradigm of "groups with social/conscious power who appear to achieve easy victories in obvious social contexts are the overdog" is flawed. Activists and universities have lots of social/conscious power, but social/conscious power is the booby prize and even in cases where it looks like it has had an effect, it has very likely just happened to fortuitously coincide with social/technological forces that changed things at the same time [again, citation needed]. If correct this observation would make a lot of reactionary thought, which focuses on activists and

universities and their ilk having too much power, kind of misguided.

Arguments About Male Violence Prove Too Much

[CONTENT WARNING: rape, violent crime, racism]

[EDIT HISTORY: This piece was widely circulated and critiqued after first being published, and I received a lot of feedback, some of which was good and some of which was bad. I have entirely rewritten the piece to try to respond to some of the complaints, especially those in the comments and those raised here. The original version of this post, without which some of the reactions and complaints will not make sense, is available here. The bottom of that post gives more information on particular edits. Thanks to everyone who gave helpful feedback both positive and negative.]

From this article:

> When the odds of being assaulted are 25 percent, something is dangerous. If any other activity or object presented the same odds of injury or death, then a revolution would be ignited against it. If one-fourth of Americans faced armed robbery in their lifetime, then you'd better believe armed robbery would be a major national issue covered everywhere in the media, and it would be right up there alongside the economy and national defense in the presidential debates.
>
> We're all willing to make a strong, concerted efforts to see that safety is followed in cases like these, with no margin allowed for error. It's a shocking contrast to how we deal with women's safety from the men who harm them.

> It makes me wonder, what if men were declared as a public safety hazard?
>
> Could you imagine if they were recalled? Pulled off the street? “Sorry sir, you’ll have to come with me; we’ve had reports that men have been raping, beating, and killing women, and we can’t take the risk that you will, too.” Yes, it’s a ridiculous idea. But men are way more dangerous than Tylenol [which was recalled for being dangerous].
>
> You may also say, “There are plenty of men out there who don’t abuse or sexually assault women — what about them?” I say: Well, what about them?

I can’t quote the whole thing, but you should probably read it if you want a clearer picture of what’s being talked about.

So when I read this article, I feel a couple things. I feel sad about the high prevalence of rape and domestic violence, of course. But I already knew that one. I also feel other emotions. As a man who hasn’t done any of these things, I feel kind of scared and singled out and unfairly guilt-by-associationed.

I know this isn’t the first time this has happened. Articles That Tar All Men With The Same Brush are pretty common, followed by Men Who Get Offended, followed by People Telling Them They Are Wrong To Be Offended Because The Problem Of Rape Is Much More Important Than The Problem Of People Getting Offended.

But, well, I *still* feel unfairly guilt-by-associationed. So here’s an intuition pump to try to communicate why.

A Visit To Racist Dystopia

Suppose you woke up one morning and started hearing public service announcements on your radio: “Black people are

defective! Black people are a public safety hazard! Black people commit lots of violent crimes! The police should just arrest all black people, because they're too dangerous to allow on the streets!"

You do some investigation and find that it's just a small fraction of the population that believes this – maybe 10% – and they're not immediately advocating any actual consequences or policy toward black people. So that's good. But you keep hearing this same message. All your favorite blogs publish a steady stream of wildly popular articles trying to "helpfully explain" to black people why all white people are justified in fearing them. Any black person in college has to walk past posters every day listing their name and reading THIS PERSON IS A POTENTIAL MUGGER – and when they complain, they are met with indifference and an administration claiming it is merely a helpful way to raise awareness of black violence.

Hopefully you, like me, would be horrified by this state of affairs. Although certainly crime is a problem in many places, and although crime is worse in poor neighborhoods which are also disproportionately minority, this is an offensive and unproductive way of thinking about the problem.

The Value Of The Analogy

If we had to specify our exact complaints against Racist Dystopia, I think there would be at least three good ones. First, we would want to protest that only a tiny percent of black people are guilty of violence. Second, we would want to protest that people of all races are capable of violence, and that the existing campaign unfairly portrays it as solely a black problem. Third, even apart from those two complaints and even assuming raising awareness of racial violence is

something we want to do, there would be ways of sending that message that encourage stereotyping and sweeping judgments and ways of discouraging them, and the current campaign seems specifically intended to promote the former.

Let's start with the first complaint: only a tiny percent of black people are guilty of violence. How tiny? Statisticians project that about 30% of black men will go to prison sometime in their lives. Somewhere around 30-40% of prisoners are serving time for violent offenses, so if we combine those two numbers (something which requires a few assumptions, but I can't find the statistic directly so it will have to do) we get that about 10% of black men will go to prison for violent crime sometime in their lives. If we assume that black women do not go to prison (I can't find good data on this), then about 5% of black people will go to prison for violent crime at some point.

This number is very similar to another number: according to an article from the early 2000s in a feminist blog, about 4.5% of men are estimated to be rapists.

Our second complaint is that violence is a problem committed by people of all races: most notably to the public service campaign, it is committed by white people as well. But as noted before, violence concentrates disproportionately among poor populations, these are disproportionately likely to be minorities, and the effects scale up. America is about 50% men/50% female. Suppose that America were 50% black/50% white. We know that black people currently commit homicide at a rate 7.5x greater than white people, so in this hypothetical society – in the implausible case where nothing changed about neighborhoods or poverty or income gap – 88% of murders would be committed by black people. Murder is an unusually good statistic because almost all murders are investigated and so there's a low chance that much of the differential is due to

racist policing, but the numbers are about the same for other violent crime. For example, in New York City, which is approximately 50% white, 25% black, and 25% other, 78% of all shootings are black compared to 2.5% white. If we extrapolate New York City into a hypothetical 50% black/50% white society, we find that the black half would commit about 97% of the shootings and the white half about 3%. Let's average these two statistics and say that in our hypothetical society where race works like gender, 95% of violent crime would be black.

And once again, these numbers are in the ballpark of male/female rape statistics. What percent of rapes are committed by men? This is very hard to determine, because rape by women is almost never reported (victim is too embarrassed) and almost never prosecuted (people just laugh and say they bet the guy liked it). I have seen claims from 99% male (which seems very high) to 75% male (which seems very low). I do not think that 95% of rapists being men to 5% women is an impossible number. Other forms of violence are even less male-dominated; male-initiated and female-initiated domestic violence seem to be about equal.

The last complaint we might have against Racist Dystopia is not statistical but moral. *Even if* it was socially necessary to raise awareness of violent crime, and *even if* everyone was so racist that the decision to fixate on black crime had already been made, there are ways to do it that are super-awful and ways to do it that are only kinda-awful. I think the most important criteria for landing a campaign in the coveted second category would be:

– Absolute avoidance of any claim or implication that the problem is with *all* minorities and extreme and frequent

repetition of the fact that the *overwhelming* majority of minorities are non-violent.

– A focus on the fact that white people can commit crimes as well, made *at least* proportional to the amount of crime they actually commit, and if possible even a little more so in order to hammer home the message that crime is not a racial problem.

– A focus on additional reasons why you need not be terrified of every single black person you met. For example, if the majority of muggings tended to occur in a few very specific situations, like after both the mugger and the victim had been drinking, or after the mugger/victim had accepted an invitation to go to the victim/mugger's house, that would have a pretty big effect on whether or not you should live in fear of your random black co-worker or college professor.

My thesis is that we should make these same complaints against efforts that try to tar all men as rapists.

How To Use The Analogy

There's a possible fourth complaint here in which the two situations are *not* similar: black people are a really oppressed group. The more one spreads fear and stereotypes about them, the more likely it is you awaken someone's latent racism and cause damage disproportional to the "limited" fear and stereotypes you were trying to convey. For example, traditionally people have been pretty quick to engage in hate crimes against black people based partly on stereotypes exactly like the one mentioned above.

One answer to this objection might be that some men are murdered, sometimes horribly because of the climate of fear around male rape. But since I've tried to stick to statistics thus far, it would be dishonest to claim that this happens in

anywhere near the numbers that would be necessary to make the analogy stick.

A better objection might be that the issues of disprivilege and oppression, while important, are *not necessary* to make the Racist Dystopia horrible. Imagine a world in which we somehow magically prevented any white people from having their opinions influenced by the public service campaign vilifying black people. They somehow continue at exactly the same level of racism they had before, and the "only" problem with the campaign is that black people have to listen to themselves be attacked all the time and get "educated" on the importance of crime prevention if they complain. This world would be *less bad* than the world in which they also had to deal with the additional racism, but it wouldn't be a walk in the park either.

Even removing the racism angle, the Dystopia above is still bad just because of the pain and dehumanization it causes people to have to read about their group in terms of some evil Other who must be a threat to all right-thinking people.

Men are blessed with many positive role models, but the divide between social and structural power is worth taking into account here, and the sort of men who are exposed to feminist articles like the one above (exactly the sort of men who are in the best position to help women!) have to take in quite a bit of information. For example, this morning when I checked Facebook I was helpfully suggested links to the "all men should be taken off the streets" article above, a blog called "Creepy White Guys", an OKCupid app that claimed to be able to tag people you looked at as "likely rapists" based on a sketchy machine learning claim about their profiles, and a link to an article called "Straight White Male: The Lowest Difficulty Level There Is". My RSS reader then directed me to

my favorite blog, which had suspended its usual discussion of abstruse philosophy to host an article called “Submissions on Misogyny”, which included bits like where someone talked about her boyfriend raping and physically assaulting her and then claimed “You might think my ex was a sociopath, but no — he’s a normal male”.

This was, more or less, a typical day for a somewhat liberal guy on the Internet. So this idea that men never have to hear anyone speaking against them and these sorts of “all men are dangerous and defective” articles are just an unexpected breath of fresh air no longer track reality, if they ever did. And if you think that a man can’t possibly be hurt by seeing people insult and belittle his gender (which, no offense, is actually a kind of patriarchalist opinion right there), all I can say is that my personal experience begs to differ.

This is *not* a demand that people stop talking about rape, or even that they stop talking about the importance of preparedness for rape! I know my feelings aren’t as important as *that*!

But it is a polite request that you follow the three suggestions I would have made to the Racist Dystopians above:

– Absolute avoidance of any claim or implication that the problem is with *all* men and extreme and frequent repetition of the fact that the *overwhelming* majority of men are non-violent. 95% of men have never raped anyone and would be horrified by the idea. Insofar as you give yourself the task of “warning women what to expect from men”, one thing they should expect is to start with a 95% probability a given man is not a rapist, and then start adjusting from there based on evidence.

– A focus on the fact that women can commit rape and gendered violence as well, made *at least* proportional to the amount of rape/violence they actually commit, and if possible even a little more so in order to hammer home the message that what we're against is rape itself and not This One Hated Out-Group.

– A focus on additional reasons why you need not be terrified of every single man you meet. This is something that feminists already do very well, in that they help explain what the warning signs of rapists are and what situations and requests are red flags for someone who might try to rape you, but this tends to be forgotten in articles like the one above which focus on scare-mongering the idea that *it could be anybody!* While it's true that it could be anybody, it's also important to keep in mind that it is somewhat more likely to be some people than others.

It's easy to see why doing this would benefit men, and I admit that's mostly why I'm writing this, but I can think of many reasons this would be good for women as well.

First, encouraging a woman to fear and distrust all men is probably not a useful strategy in a society that's fifty percent male, especially if that woman happens to be heterosexual. A strategy of "be aware of this possibility and of the warning signs, but also that most of the people you interact with are nice and trustworthy" is probably both psychologically and socially more healthy.

Second, women have a vested interest in fighting sexism and sexist stereotypes. Sexism is basically a flawed cognitive algorithm. It's the tendency to think "I can think of a bunch of people of this sex who do X, therefore I'm just going to classify that sex as X-doers and promote that idea to society."

A big part of fighting sexism is discouraging this process. Saying “No, you can’t just say that because some women like cooking in this society, cooking is a Thing Inherent About Women. Further, we can’t even create a climate where women are constantly portrayed as cooking and doing things relating to cooking, because that’s going to make non-culinary women feel bad.” And if you laboriously train people out of this habit of thought, and then say “But definitely do this for men, they don’t count because they’re privileged oppressors”, it’s not going to work for women either. You’re creating a natural Stroop effect where people have to keep conflicting category-based rules in their head.

Third, and most speculatively, I’m kind of worried that this sort of stereotype actually promotes rape. There was a very interesting study where researchers interviewed some people about their relationships, and then told half the subjects (at random) that they had been commendably faithful, and the other half that their actions suggested they were of an unfaithful personality type and their mental infidelity might destroy their relationship. Then they asked the subjects for their opinions on infidelity. The subjects who had been told they were commendably faithful told them fidelity was extremely important to them; the subjects who had been (randomly!) told they were unfaithful told them that fidelity wasn’t important to them and that infidelity wasn’t a big deal anyway.

The researchers theorized that this was a process called “cognitive dissonance”. Most people like themselves and want to continue to like themselves. If they are told that they, or their group, has a particular flaw, then instead of ceasing to like themselves it may be easier to just decide that flaw is not

such a big deal and they can have it while continuing to be the awesome people they secretly know they are.

If rape is portrayed as inherent to men in some way, men have two mental choices. They can think “darn, I guess I got the evil gender.” Or they can think “well, if *my* gender does it, I guess it isn’t so bad”. Psychology suggests there will be at least some small tendency to react with the second.

I don’t know how important this effect is, but given that just stating the case for rape awareness respectfully and non-prejudicially is an easy and desirable solution anyway, I don’t see why one should take any chances.

Disclaimers That Should Not Be Necessary, But Are

I am not trying to compare the experience of men to the experience of black people in any way other than the very simple numerical comparisons listed.

I am not actually saying that black people should be seen as criminals, I am using this as an example of a bad argument in order to show that tarring all men with the crime of rape is also a bad argument.

I am not claiming straight white men are not privileged or do not have things easier than other groups.

I am not apologizing for rape or claiming it is anything other than really bad. I am not denying women the right to avoid or fear men if that is what makes them comfortable.

Social Justice for the Highly-Demanding-of-Rigor

My last two posts have led to a lot of anti-feminist activists getting linked to my blog, so this would be a more hilarious time than usual to write the next post in my series of arguments against Reactionary politics – about why fighting racism and sexism is necessary and important.

The Reactionary argument, as I understand it, is twofold.

First, that social justice advocates irresponsibly take some undesirable outcome in minority groups, like poverty, and then assume it is the result of racism or sexism without considering other possible explanations.

Second, that a disproportionate amount of time and energy is spent worrying about this, in a way that can only be explained through wasteful signaling cascades.

My counterargument is that although the first argument is true a depressingly large amount of the time, some people do more rigorous work and get the same result – that poor outcomes for minority groups are caused in large part by racism and sexism. And second, that these poor outcomes for minority groups are a major problem even by objective quantifiable standards.

Controlled Experiments On Prejudice

The most fun experiments on prejudice are Implicit Association Tests, which test people's reaction times in linking together different concepts. If these concepts are socially important (for example, the concepts "white person", "black person", "good", and "evil") it can test how closely two different concepts are linked. The best way to get a feel for this is to take one yourself.

88% of white Americans (and 48% of black Americans!) show an implicit racial preference for whites on this test. How does that translate into the real world?

Some of the most interesting controlled experiments are detailed in an early ’90s review article in the Journal of Black Political Economy. A consortium of interested parties such as the Fair Employment Committee teamed up with recent university graduates. They laboriously paired white and black graduates by similar attractiveness, well-spokenness, age, gender, and qualifications (in some cases, the qualifications were faked to be as similar as possible), then sent them off job-hunting to the same companies.

In these sorts of experiments, 48% of white testers and 40% of black testers received interviews, a small and in fact nonsignificant difference. *However*, 47% of interviewed whites were offered jobs, compared to only 11% of interviewed blacks – a *gigantic* difference. Multiplying these two numbers together, we find that 23% of whites and 4% of blacks involved in the experiment got jobs – a difference of almost 6x. The whites also got a few other minor advantages – very slightly higher wages and slightly more likelihood of being informed of other open positions at the company.

Another good review article is in the Annals of the American Academy of Political and Social Sciences. It lists a few similar studies. In one such study, researchers, instead of training real applicants, send off fake resumes with extremely white-sounding or extremely black-sounding names; they find employers respond to the white-sounding names about 50% more often. But it also has some studies of in-person interview similar to the ones above. These studies, which are from the mid-2000s rather than the early 1990s, feature white:black success ratios of anywhere from 1.5x to 5x.

Along with labor discrimination, it's harder for minorities to buy things. For example, when trying to buy a car, black men were asked to pay on average $1100 more than attribute-paired white men. Interestingly enough, black car salesmen, and black owned car dealerships, displayed this pattern to exactly the same degree as white-owned institutions.

The situation is roughtly similar in housing. In an experiment where researchers responded to Craigslist notices advertising apartments in Toronto, using names of various ethnicities, they found that Caucasian experimenters confused relative risk with odds ratios 100% of the t…ahem, sorry, they found black people experienced housing discrimination 5% of the time and Muslims 12% of the time, usually in the form of not receiving a response even when the white person was simultaneously invited to come on over. A similar study in Houston found an astronomical 80% discrimination rate for blacks, so either Houston is much worse than Toronto, someone's not doing their studies properly, or I'm misinterpreting something.

Other experiments along the same lines include a cute little bus experiment in Sydney where someone got on bus, their travel card didn't work, and they asked the driver to let them ride anyway. For whites (and Asians) it worked about 72% of the time; for Indians, about 50%, and for blacks, 36%. In a later survey, bus drivers (who were unaware the experiment was going on) claimed they would prefer to help black people over white people. Interestingly enough, although black bus drivers were a bit nicer to blacks than white bus drivers, they *still* let whites and Asians on more often.

We find much the same pattern with men and women. A famous study a few months ago found that faculty offered a female grad student a 12% lower salary than an identical male

grad student (again interestingly, female faculty were *more* biased against female grad students than male faculty were).

A less perfect but more natural experiment is switching from an open application procedure where applicants' genders are obvious to a blind procedure in which genders are unclear. If the percent of women hired increases (and perhaps if no similar increase is seen in competitors that don't change procedure at the same time) this implies the institution was being unfairly biased before. ~~When such a test was performed by the Journal of Behavioral Ecology starting in 2001, the percent of articles by female authors went up from about 29% to about 37%, about a 30% increase~~ **.[EDIT: This has since been found to be false]** Symphony orchestras are another infamous example, and studies show that the switch from open to blind auditions explains between half and a third of the recent quintupling of the percent women in symphonies over the past thirty years.

What do these show and not show? They show that, even controlling for all other factors like different preferences, different negotiating strategies, different educational backgrounds, et cetera there is a large difference in the opportunities of minority and majority groups due solely to discrimination. This difference seems large enough to explain the proportion of the income gaps that people say it explains (usually around half of the gap for each minority group) and to give minorities large amounts of trouble throughout the rest of their lives.

One thing it does not show is that racism is just about straight white men being evil. Minorities seem just as willing to screw other minorities over and discriminate in favor of white men as the white men themselves are. A better model would be that ideas of certain races and genders being superior seem to

percolate into people's consciousnesses, regardless of what race those people themselves are, and shape their actions whether they mean for them to or not.

Economic Costs of Discrimination

A beautiful experiment by Gwartney and Haworth noticed that baseball formed a natural experiment about the costs of discrimination. During the post-Jackie-Robinson 1950s, some teams had integrated but others had not and remained white-only. G&H wondered whether this affected performance. They found that in fact the five teams quickest to integrate black players were five out of the six top performers in the league, and that every additional black player on a team resulted in an addition 3.75 wins. This was partially because black players outperformed whites on average, but also because there was more low-hanging fruit in the form of black talent which could be employed more cheaply.

What is true for baseball teams is probably also true for other companies, but harder to quantify. For such a popular field, I cannot for the life of me find any attempt to quantify the economic costs of racism. There seem to be some people in Australia working on it, but they have yet to publish any results. So let's make some up (this, uh, ends the demanding-of-rigor part of this post).

One way to do this is to take people's estimates of the purely-discriminatory pay gap for different groups – that is, how much less they earn than straight white men when all other factors anyone can think of (like education level, IQ, height, region of residence, whatever) are adjusted away. Then multiply this by the number of people in that group and their average wage, and we get part of the cost of racism per year.

One of the review articles above suggests the black pay gap is 15%; others suggest numbers around 10% for women. Asians and gays make a bit more than straight white men, and although Latinos make much less no one has bothered adjusting for confounders so I can't include them.

Anyway, when I add all that up, I get $374 billion.

(one might argue that the companies these people work for *gain* this money as profit by paying employees less, so it all evens out. I don't think that works. In at least some cases, the lower pay must be because they have lower-level jobs than their white male counterparts. But since we already agreed they have the same skills as their white male counterparts, this suggests their skills aren't being used fully, which means the cost really is to the economy and not just to them. I have no idea whether this argument works in real life. Like I said, the highly-demanding-of-rigor probably should have stopped with the first half of this post)

Another claim is that companies lose $64 billion dollars to discrimination-related turnover yearly. The number is generated by taking the cost of replacing a lost employee with results of surveys about how many people leave due to discrimination or hostility at their former place of employment.

Suppose we arbitrarily and implausibly stop here because we're tired. We've found costs of US racism equal to at least $438 billion per year.

That's *about* the annual budget of the US military.

Note what it doesn't include. It doesn't include any non-monetary costs like people being unhappy. It doesn't include the costs to the prison system of overprosecuting minorities. It doesn't include the costs to the health system of minorities

getting worse preventative health. It doesn't include the amount the government spends fighting racism, or the amount people have to pay out in racism lawsuits. It doesn't include people who are unemployed because of racism, because I took the data from the employment records. It doesn't include any of the income gap due to racism anywhere other than at job – for example, racism that affects how much education people of different races end up with, or racism the person's parent suffers that then screws up their families for several generations. It doesn't even include Latinos because I couldn't find any good numbers about them.

It seems *very* unlikely to me that the actual costs are less than $1 trillion/year in the US alone. But let's stick with the $438 billion figure.

(another way to look at this is that these arbitrarily-stopped at costs of racism/sexism are about $2-3K per minority group member in the US, counting women as a "minority group". This seems broadly reasonable, and is in fact still way less than the observed non-adjusted income gap)

What other things cost $438 billion dollars? According to the American Cancer Society, cancer costs $200 billion/year Heart disease costs $100 billion. Adding up all of the easy-to-calculate costs of 9/11 on this page, I get about $250 billion.

So in terms of purely economic, not-even-worrying-about-human-beings costs, the costs of racism and sexism that can be pretty plausibly attributed to discrimination alone are equivalent to about heart disease plus cancer plus half a 9/11 or so per year.

One good thing about the size of this number means that small successes in fighting racism and sexism are extremely valuable. For example, decreasing racism/sexism by 1% is a

$4.4 billion gain per year to the economy, which is about equal to Facebook's 2011 annual revenue.

Effectiveness

Now none of this is meant to claim that the marginal blog on Tumblr complaining about the patriarchy has positive expected value or is anything other than a massive waste of everyone's time.

But there are aspects of the social justice movement interested in testing what works and doing it.

As of yet, I don't think most of them are aware of the pitfalls in claiming successful interventions – all of these "We found our intervention decreases expressions of prejudice on a seven point scale two weeks later!" things sound suspiciously like "Our drug increases 'good cholesterol' after three days, and we didn't bother to check whether it actually prevents heart attacks but seriously how could it not?"

But we can't blame them for their failure to be more rigorous than the hard sciences, and besides one day they might wise up.

With Implicit Association Tests, Ultimatum/Dictator games, and the like, I think there is a decent toolkit for people who want to wise up and seriously analyze anti-racism and anti-sexism strategies, and I bet when they are tested further some of the ones in that document will turn out to work longer-term. Maybe they could decrease racism by 1% a year and save us $4 billion or so.

The fact that racism in the sense of simple prejudice is a real problem that accounts for much of the disadvantage of minorities; that it has a huge negative effect even when you try to measure it objectively; and that it can be fought – seem to

take some of the wind out of the Reactionary argument against social justice.

Against Bravery Debates

One of the things I was most criticized for on my old blog – and upon reflection, criticized for fairly – was my propensity to engage in bravery debates.

There's a tradition on Reddit that when somebody repeats some cliche in a tone that makes it sound like she believes she is bringing some brilliant and heretical insight – like "I know I'm going to get downvoted for this, but believe we should have *less* government waste!" – people respond "SO BRAVE" in the comments. That's what I mean by bravery debates. Discussions over who is bravely holding a nonconformist position in the face of persecution, and who is a coward defending the popular status quo and trying to silence dissenters.

These are *frickin' toxic*. I don't have a great explanation for why. It could be a status thing – saying that you're the original thinker who has cast off the Matrix of omnipresent conformity and your opponent is a sheeple (sherson?) too fearful to realize your insight. Or it could be that, as the saying goes, "everyone is fighting a hard battle", and telling someone else they've got it easy compared to you is just about the most demeaning thing you can do, especially when you're wrong.

But the possible explanations aren't the point. The point is that, empirically, starting a bravery debate is the quickest way to make sure that a conversation becomes horrible and infuriating. I'm generalizing from my own experience here, but one of the least pleasant philosophical experiences is thinking you're bravely defending an unpopular but correct position, facing the constant persecution and prejudice from your more numerous and extremely smug opponents day in

and day out without being worn-down … only to have one of your opponents offhandedly refer to how brave they are for resisting the monolithic machine that you and the rest of the unfairly-biased-toward-you culture have set up against them. You just want to scream NO YOU'RE WRONG SEFSEFILASDJO:IALJAOI:JA:O>ILFJASL:KFJ

A lot of common political terms pretty much encode bravery debates. "Political correctness", "mainstream media", "liberal media", "corporate media", "rape culture", "Big Government" or "Big Business" or "Big Anything", "patriarchy", "the climate establishment", or "the anything-anything complex". By not-at-all-a-coincidence, these also happen to be some of the terms most likely to be inflammatory and get people angry. Has there *ever* been an argument that continued being civil or productive after "political correctness" was mentioned?

The persistence of bravery debates is actually kind of weird. Shouldn't it be really really easy to figure out who's being oppressed by whom? The Spanish Inquisition had many faults, but whining about being unfairly persecuted by heretics was, as far as I know, not one of them. Can two opposing positions really be absolutely certain they are under siege?

This question immediately reminded me of my recent observation about Christians and Muslims in the media. Whenever the media says something negative about Christians, comments and blogs and forums immediately fill up with claims that the media loves picking on Christians and that no one would ever publish a similar story about Muslims for fear of being "offensive" (eg 1, 2, 3, 4, 5). And whenever the media says something negative about Muslims, comments and blogs and forums immediately fill up with claims that the media is Islamophobic and attacks Muslims any chance it gets

and they would never dare pick on a large powerful group like Christians in such a way (eg 1, 2, 3, 4, 5).

So for example, Aziz Mubaraki writes:

> There are numerous cases to judge whether there is bias against Muslims in the media, but in recent times look no further than the press coverage regarding the terrorist attack that took place in Norway not very long ago. Impartial population waited impatiently to read this act being explicitly described as a "terrorist attack" or an "act of terrorism" by the mainstream media. But never once the "Christian" label was used despite the fact that Mr. Breivik was a self-described devout Christian. Therefore the important question is: Why is it when the person responsible for the terrorist act happens to be Muslim all of a sudden the religion becomes the focus instead?

Yet israpundit.com writes:

> Big media has no qualms about boldly and repeatedly labeling the Norweigan shooter as a "Christian", even describing him as a Christian Zionist, despite no evidence that he was any kind of devout Christian whatsoever. Yet till this day the same vile liberal media will not refer to the Fort Hood jihadist as muslim or emphasize the Islamic motivation behind the shooting. Neither do government reports on the jihad attack.

So can we agree that this phenomenon of two opposing groups being equally sure they are bravely pointing out the world's bias in favor of the other is, in fact, a thing?

Because once we acknowledge it, it's not really hard to explain.

Psychologists have known about the hostile media effect for thirty years, ever since a 1982 study where they got pro-Israeli and pro-Palestinian students to watch a documentary and found that:

> On a number of objective measures, both sides found that these identical news clips were slanted in favor of the other side. Pro-Israeli students reported seeing more anti-Israel references and fewer favorable references to Israel in the news report and pro-Palestinian students reported seeing more anti-Palestinian references, and so on. Both sides said a neutral observer would have a more negative view of their side from viewing the clips, and that the media would have excused the other side where it blamed their side.

Note that this was not at all subtle. The pro-Palestinians claimed that favorable references to Israel outnumbered unfavorable references almost 2:1, but the pro-Israelis complained that unfavorable references outnumbered favorable references at a greater than 3:1 ratio ($p < .001$). Transforming a different measure mentioned earlier in the paper to a scale of 1 to 10, where 1 is completely pro-Palestine and 10 is completely pro-Israel, the average pro-Israeli rated it a 3.2, and the average pro-Palestinian rated it a 7.4. These numbers were even higher in people who claimed to know a lot about the conflict. So even when exposed to genuinely neutral information, people tend to believe the deck is stacked against them. But people aren't exposed to genuinely neutral information. In a country of 300 million people, *every single day* there is going to be an example of something hideously biased against *every single group*, and proponents of those groups have formed effective machines to publicize the most

outrageous examples in order to "confirm" their claims of bravery. I had an interesting discussion on Rebecca Hamilton's blog about the Stomp Jesus incident. You probably never heard of this, but in the conservative Christian community it was a *huge deal*; Google gives 20,500 results for the phrase "stomp Jesus" in quotation marks, including up-to-date coverage from a bunch of big conservative blogs, news outlets, and forums. I guarantee that the readers of those blogs and forums are *constantly* fed salient examples of conservatives being oppressed and persecuted. And I don't mean "can't put up ten commandments in school", I mean armed gay rights activist breaks into Family Research Council headquarters and starts shooting people for opposing homosexuality. Imagine you hear a story in this genre almost every time you open your RSS feed.

(And now consider all the stories *you* hear every day about violence and harassment against *your* people in *your* RSS feed.)

And if there aren't enough shooters, someone is saying something despicable on Twitter pretty much every minute. The genre of "we know the world is against us because of five cherry-picked quotes from Twitter" is alive, well, and shaping people's perceptions. Here's an atheist blog trawling Twitter for horrible comments blaming atheists for terrorism, and here's an article on the tweets Brad Pitt's mother got for writing an editorial supporting Romney (including such gems as "Brad Pitt's mom wrote an anti-gay pro-Romney editorial. Kill the b—-.")

Then we get into more subtle forms of selection bias. Looking at the articles above, I am totally willing to believe newspapers are more likely to blaspheme Jesus than Mohammed, and also that newspapers are more likely to call a Muslim criminal a

"terrorist" than they would a Christian criminal. Depending on your side, you can focus on one or the other of those statements and use it to prove the broader statement that "the media is biased against Christians/Muslims in favor of Muslims/Christians". Or you can focus on one part of society in particular being against you – for leftists, the corporations; for rightists, the universities – and if you exaggerate their power and use them as a proxy for society then you can say *society* is against you. Or as a last resort you can focus on only one side of the divide between social and structural power.

So it's far from a mystery how bravery debates can be so common or persistent. Or why everyone is so sure they're on the brave side. But the interesting thing is that they actually *work*.

I call your attention to two studies by Joseph Vandello et al. In the first, experimenters once again took the Israeli-Palestinian conflict but ran the experiment in the other direction. Here they presented maps that showed Palestine as the underdog (by displaying a map emphasizing a tiny Palestine surrounded by much larger Israel) or Israel as the underdog (by displaying a map emphasizing tiny Israel surrounded by a much larger Arab world including Palestine). In the "Palestinians as underdogs" condition, 55% of subjects said they supported Palestine. In the "Israelis as underdogs" condition, 75% said they supported Israel. And in the second, experimenters found subjects rated people who had been unfairly disadvantaged during a job interview as more attractive and more desirable romantic partners than people who had not been.

Baaaaasically if you get yourself perceived as the brave long-suffering underdog, people will support your cause and, as an added bonus, want to have sex with you.

And I dislike this, because bravery debates tend to be so fun and addictive that they drown out everything more substantive. Sometimes they can be acceptable stand-ins for actually having an opinion at all. I constantly get far-right blogs linking to my summary of Reactionary thought, and I hope I'm not being too unfair when I detect an occasional element of "Oh, so *that's* what our positions are!". There seem to be a *whole lot* of Reactionaries out there who are much less certain of what they believe than that they are very brave and nonconformist for believing it.

As I said before, I accept the criticism that I was too quick to start bravery debates at my old blog and am trying to cut down on them. I would also recommend that other people cut down on them. I think they probably fall into the large category of things that make people who already agree with you fist-pump and shout "Yeah! We *are* awesome rebels!" while alienating everyone who doesn't hold your position.

But what if you *are* being really brave by holding a dangerous and unpopular position? Shouldn't you get credit for that?

I guess. I propose that if you write something and, for even just a second, you think of not publishing it, because of the risk to your reputation, or your livelihood, or your family, or even your life – then go ahead and call yourself brave, and I will try to reassure you and tell you everything is going to be all right.

If you think "Not publish this? But then how would everyone know how brave I'm being? I'm going to plaster my name all over this thing so everyone knows exactly where to send the bravery-related kudos!" … then stick to the damn object-level issues.

All Debates Are Bravery Debates

"I don't practice what I preach because I'm not the kind of person I'm preaching to."
— Bob Dobbs

I.

I read Atlas Shrugged probably about a decade ago. I was impressed with its defense of capitalism, which really hammers home the reasons it's good and important on a gut level. But I was equally turned off by its promotion of selfishness as a moral ideal. I thought that was *basically* just being a jerk. After all, if there's one thing the world doesn't need (I thought) it's more selfishness.

Then I talked to a friend who told me Atlas Shrugged had changed his life. That he'd been raised in a really strict family that had told him that ever enjoying himself was selfish and made him a bad person, that he had to be working at every moment to make his family and other people happy or else let them shame him to pieces. And the revelation that it was sometimes okay to consider your own happiness gave him the strength to stand up to them and turn his life around, while still keeping the basic human instinct of helping others when he wanted to and he felt they deserved it (as, indeed, do Rand characters).

II.

The religious and the irreligious alike enjoy making fun of Reddit's r/atheism, which combines an extreme strawmanning of religious positions with childish insults and distasteful triumphalism. Recently the moderators themselves have become a bit embarrassed by it and instituted some rules

intended to tone things down, leading to some of the most impressive Internet drama I have ever seen. In its midst, some people started talking about what the old strawmanning triumphalist r/atheism meant to them (see for example here).

A lot of them were raised in religious families where they would have been disowned if they had admitted to their atheism. Some of them *were* disowned for admitting to atheism, or lost boyfriends/girlfriends, or were terrified they might go to Hell. And then they found r/atheism, and saw people making fun of religion, and insulting it, in really REALLY offensive ways. And no one was striking them down with lightning. No one was shouting them down. No one was doing much of anything at all. And to see this taboo violated in the most shocking possible way with no repercussions sort of broke the spell for them, like as long as people were behaving respectfully to religion, even respectfully disagreeing, it still had this aura of invincibility about it, but if some perfectly normal person can post a a stupid comic where Jesus has gay sex with Mohammed, then there's this whole other world out there where religion holds no power.

Gilbert tells the story of how when, as a young Christian struggling with doubt, he would read r/atheism to remind himself that atheists could be pretty awful. r/atheism is doing a bad job at being the sort of people who can convert Gilbert, and the new mods' policy of "you should have more civil and intellectual discussions" might work better on him. I think it would work better on me too.

But there is – previously unappreciated by me – a large population of people for whom really dumb offensive strawmannish memes are *exactly what they need*.

III.

My last night in Berkeley, I went to a CFAR party where someone (can't remember who) was talking about his experiences in Landmark Forum, a self-improvement workshop. He said their *modus operandi* was to get people to take responsibility for the outcome of their actions. His example was an office worker who always did substandard work, and was always making excuses like "My boss doesn't s support me" or "My computer system isn't good enough" or "My coworkers aren't pulling their fair share." Landmark says those kinds of excuses are what's keeping you back. And they taught (again, according to this one person) that the solution was to treat everything that happens in your life as your responsiblity – no excuses, just "it was my fault" or "it's to my credit".

Which is probably a really good idea for this guy in his one job. But someone else at the party pointed out that there are situations where this heuristic is horrible. Like if you're a teenager trying to cope with the trauma of your parents' divorce.

(Around the same time, I saw the same idea expressed as a Rationality Quote on Less Wrong: "Bad things don't happen to you because you're unlucky. Bad things happen to you because you're a dumbass." Eliezer went a good deal of the way to correcting it by rephrasing to "Single bad things happen to you at random. Iterated bad things happen to you because you're a dumbass." But I would go further and add "Or because you're a minority. Or because you live in an awful place, like the ghetto, or North Korea. Or because there's a war going on. Or because you have a disease, either somatic or psychiatric. Or because of any of the thousand and one other reasons why you might consistently have bad things happen to you that aren't your fault.")

And only a few days after the party, I was reading a book on therapy which contained the phrase (I copied it down to make sure I got it right) "Don't be so hard on yourself. No one else is as hard on yourself as you are. You are your own worst critic."

Notice that this encodes the *exact opposite assumption*. Landmark claims its members are biased against ever thinking ill of themselves, even when they deserve it. The therapy book claims that patients are biased towards always thinking ill of themselves, even when they do deserve it.

And you know, both claims are probably spot on. There are definitely people who are too hard on themselves. Ozy (previously on my blogroll) has done an amazing job of getting me and many other people inclined towards skepticism about feminist and transgender issues, engaging with us, and gradually convincing us to be more respectful and aware through sheer kindness and willingness to engage people reasonably on every part of the political spectrum. Two days ago some people on Twitter – who were angry Ozy said one need not boycott everything Orson Scott Card has ever written just because he's against gay marriage – told Ozy ze wasn't a real transgender person and suggested lots of people secretly disliked zir. And instead of doing what I would do and telling the trolls to go to hell, Ozy freaked out and worried ze was doing everything wrong and decided to delete everything ze had ever written online. I *know* Ozy is zir own worst critic and if that therapy book was aimed at people like zir, it was entirely correct to say what it said.

On the other hand, I look at people like Amy's Baking Company, who are obviously terrible people, who get a high-status professional chef as well as thousands of random joes informing them of exactly what they are doing wrong, who are

so clearly in the wrong that it seems impossible not to realize it – and who then go on to attribute the negativity to a "conspiracy" against them and deny any wrongdoing. They could probably use some Landmark.

IV.

In a recent essay, Against Bravery Debates, I think I underestimated an important reason why some debates *have to* be bravery debates.

Suppose there are two sides to an issue. Be more or less selfish. Post more or less offensive atheist memes. Be more or less willing to blame and criticize yourself.

There are some people who need to hear both sides of the issue. Some people really need to hear the advice "It's okay to be selfish sometimes!" Other people really need to hear the advice "You are being way too selfish and it's not okay."

It's really hard to target advice at exactly the people who need it. You can't go around giving everyone surveys to see how selfish they are, and give half of them *Atlas Shrugged* and half of them the collected works of Peter Singer. You can't even write really complicated books on how to tell whether you need more or less selfishness in your life – they're not going to be as buyable, as readable, *or* as memorable as *Atlas Shrugged*. To a first approximation, all you can do is saturate society with pro-selfishness or anti-selfishness messages, and realize you'll be hurting a select few people while helping the majority.

But in this case, it makes a really big deal what the majority actually is.

Suppose an Objectivist argues "Our culture has become too self-sacrificing! Everyone is told their entire life that the only

purpose of living is to work for other people. As a result, people are miserable and no one is allowed to enjoy themselves at all." If they're right, then helping spread Objectivism is probably a good idea – it will help these legions of poor insufficiently-selfish people, but there will be very few too-selfish-already people who will be screwed up by the advice.

But suppose Peter Singer argues "We live in a culture of selfishness! Everyone is always told to look out for number one, and the poor are completely neglected!" Well, then we want to give everyone the collected works of Peter Singer so we can solve this problem, and we don't have to worry about accidentally traumatizing the poor self-sacrificing people more, because we've already agreed there aren't very many of these at all.

It's much easier to be charitable in political debates when you view the two participants as coming from two different cultures that err on opposite sides, each trying to propose advice that would help their own culture, each being tragically unaware that the other culture exists.

A lot of the time this happens when one person is from a dysfunctional community and suggesting very strong measures against some problem the community faces, and the other person is from a functional community and thinks the first person is being extreme, fanatical or persecutory.

This happens a lot among, once again, atheists. One guy is like "WE NEED TO DESTROY RELIGION IT CORRUPTS EVERYTHING IT TOUCHES ANYONE WHO MAKES ANY COMPROMISES WITH IT IS A TRAITOR KILL KILL KILL." And the other guy is like "Hello? Religion may not be literally true, but it usually just makes people feel more

comfortable and inspires them to do nice things and we don't want to look like huge jerks here." Usually the first guy was raised Jehovah's Witness and the second guy was raised Moralistic Therapeutic Deist.

But I've also sometimes had this issue when I talk to feminists. They're like "Guys need to be more concerned about women's boundaries, and women need to be willing to shame and embarrass guys who hit on them inappropriately." And maybe *they* spent high school hanging out with bros on the football team who thought asking women's consent was a boring technicality, and *I* spent high school hanging out entirely with extremely considerate but very shy geeks who spent their teenage years in a state of nightmarish loneliness and depression because they were too scared to ask out women because the woman might try to shame and embarrass them for it.

And the big one is trust. There are so many people from extremely functional communities saying that people need to be more trusting and kind and take people at their word more often, and so many people from dysfunctional communities saying that's not how it works. Both are no doubt backed by ample advice from their own lives.

A blog like this one probably should promote the opinions and advice most likely to be underrepresented in the blog-reading populace (which is *totally different* from the populace at large). But this might convince "thought leaders", who then use it to inspire change in the populace at large, which will probably be in the wrong direction. I think most of my friends are too leftist but society as a whole is too rightist – should I spread leftist or rightist memes among my friends?

I feel pretty okay about both being sort of a libertarian and writing an essay arguing against libertarianism, because the world generally isn't libertarian enough but the sorts of people who read long online political essays generally are way more libertarian than can possibly be healthy.

A Comment I Posted on "What Would JT Do?"

Last week JT posted An Open Letter To The Defenders Of Phil Robertson, which bothered me enough that I posted the following:

> So I'm the person you are insisting doesn't exist – a completely pro-gay atheist who voted against Proposition 8 and thinks supporting gay marriage is a no-brainer, but who is also kind of horrified at Phil Robertson being fired for his comments.
>
> You are 100% correct that freedom-of-speech only binds the government and does not constrain private actors from punishing people whose speech they don't like.
>
> But let's compare and contrast. Freedom of religion *also* only binds the government and does not constraint private actors from punishing people whose religion they don't like. If someone wants to picket a mosque while waving signs about how all Muslims are dirty terrorists who are going to Hell, Constitutional freedom of religion is a-ok with that. Heck, Constitutional freedom-of-religion is okay with Christian-owned businesses refusing to hire atheist employees or serve atheist customers – it's only more recent anti-discrimination laws that prevent that.
>
> Point is, there's a big gap between "constitutional freedom of religion" and "the level of religious tolerance that is necessary to have a remotely civil society." Some of that gap can be filled in by laws, but a lot of it can't be. It's supposed to be filled in by basic human decency and

understanding of the principles that made freedom of religion a good idea to begin with.

I think the same is true of freedom of speech. Constitutional freedom-of-speech is a necessary but not sufficient condition to have a "marketplace of ideas" and avoid de facto censorship. But people also have to understand that the correct response to "idea I disagree with" is "counterargument", not "find some way to punish or financially ruin the person who expresses it." If you respond with counterargument, then there's a debate and eventually the people with better ideas win (as is very clearly happening right now with gay marriage). If there's a norm of trying to punish the people with opposing views, then it doesn't really matter whether you're doing it with threats of political oppression, of financial ruin, or of social ostracism, the end result is the same – the group with the most money and popularity wins, any disagreeing ideas never get expressed.

Atheists may one day be the group with the most money and popularity, but that day isn't today and right now it's neither moral nor in our self-interest to encourage using greater resources to steamroll opponents. It's certainly not in gay people's self-interest either. Why shouldn't companies owned by Christians fire all gay people on the grounds that they are promoting sin? Right now it's bcause we have a mutual truce in which we agree businesses should employ people based on their skills and merit rather than to reward their political allies and punish their political opponents. Once you undermine that, gay people are in a pretty precarious position.

So I would turn your own hypothetical scenario in Part 2 of your post back on you. Suppose Robertson had indeed,

> been a gay rights supporter – or a gay person! – who said on national news he thought everyone should stand up for gay rights. But his company was going for the fundie demographic and decided to fire him for his statement. Would you be so quick to attack everyone who was disappointed in this action, so eager to stand up for the right of companies to fire anyone they disagree with?
>
> I'm an atheist blogger and I work at a Catholic hospital. Employer tolerance for dissenting opinions is *personal* for me. I'm disappointed in the tone of this post and I hope you reconsider.

I can't tell how many other people have made similar points because none of the three browsers on my computer can successfully load Patheos' nightmarish comment system more than once in a blue moon. But I hope some other Patheos atheists are saying the same. And I have huge respect for the few voices on the lefty blogosphere, like Ampersand, who have spoken out in favor of restraint.

CORRECTION: Mr. Robertson was suspended rather than fired, and has since been reinstated.

We Are All MsScribe

AskReddit asked recently: If you could only give an alien one thing to help them understand the human race, what would you give them?

At the time I had no good answer. Now I do. I would give them Charlotte Lennox's write-up of how MsScribe took over Harry Potter fandom (warning: super-long but super-worth-it).

Ozy informs me that everyone else in the world read this story five years ago. Maybe I am hopelessly behind the times? Maybe all my blog readers are intimately familiar with it?

If not, read it. Read it like an anthropological text. Read it like you would a study of the Yanomamo. No, read it even better than that. Read it like you would a study of the Yanomamo if you knew that, statistically, some of your friends and co-workers covertly become Yanomamo after getting home every evening.

I hesitate to summarize it, because people will read my summary and ignore the much superior original. I would not recommend that. But if you insist on skipping the (admittedly super-long) link above, here is what happens:

> In the early 2000s, Harry Potter fanfiction authors and readers get embroiled in an apocalyptic feud between people who think that Harry should be in a relationship with Ginny vs. people who think Harry should be in a relationship with Hermione. This devolves from debate to personal attacks to real world stalking and harassment to legal cases to them splitting the community into different sites that pretty much refuse to talk to each other and ban stories with their nonpreferred relationship.

These sites then sort themselves out into a status hierarchy with a few people called Big Name Fans at the top and everyone else competing to get their attention and affection, whether by praising them slavishly or by striking out in particularly cruel ways at people in the "enemy" relationship community.

A young woman named MsScribe joins the Harry/Hermione community. She proceeds to make herself popular and famous by use of sock-puppet accounts (a sockpuppet is when someone uses multiple internet nicknames to pretend to be multiple different people) that all praise her and talk about how great she is. Then she moves on to racist and sexist sockpuppet accounts who launch lots of slurs at her, so that everyone feels very sorry for her.

At the height of her power, she controls a small army of religious trolls who go around talking about the sinfulness of Harry Potter fanfiction authors and *especially* MsScribe and how much they hate gay people. All of these trolls drop hints about how they are supported by the Harry/Ginny community, and MsScribe leads the campaign to paint everyone who wants Harry and Ginny to be in a relationship as vile bigots and/or Christians. She classily cements her position by convincing everyone to call them "cockroaches" and post pictures of cockroaches whenever they make comments.

Throughout all this, a bunch of people are coming up with ironclad evidence that she is the one behind all of this (this is the Internet! They can just trace IPs!) Throughout all of it, MsScribe makes increasingly implausible denials. And throughout all of it, everyone supports MsScribe and ridicules her accusers. Because

> really, do you want to be on the side of a confirmed popular person, or a bunch of confirmed suspected racists whom we know are racist because they deny racism *which is exactly what we would expect racists to do?*
>
> MsScribe writes negatively about a fan with cancer asking for money, and her comments get interpreted as being needlessly cruel to a cancer patient. Her popularity drops and everyone takes a second look at the evidence and realizes hey, she was obviously manipulating everyone all along. There is slight sheepishness but few apologies, because hey, we honestly thought the people we were bullying were unpopular.

MsScribe later ended up switching from Harry Potter fandom to blogging about social justice issues, which does not surprise me one bit. But let me do some social justice blogging of my own.

A lot of the comments I have seen discussing the issue say "Yeah, teenage girls will be teenage girls".

Two responses seem relevant. First, quite a few of the people involved seem to have been in their late twenties or early thirties.

But second and more important, I am a guy and this story speaks to me because it is *eerily* similar to the story of my online life with a bunch of other guys when I was between about ages fifteen and twenty-two.

I've mentioned before how I spent long portions of my life in the interactive geofiction/"micronation" community. And because of the innate urge for self-presentation, I emphasized the part where we create amazing grand-scale fictional universes in which we enact epic battles and build civilizations

from the ground up. And not the part where we behave like ridiculous little children having a hissy fit.

The first constructed country I was ever in, another guy named John from my school comes in and says that I am a bad leader and abusing my power. Because my online handle at the time was Giant_Squid314, he classily nicknames me “Squitler” and leads a bunch of his supporters to make “Squitler” related comments at everything I do. Then he and his friends secede to start their own country, named after a Red Hot Chili Peppers album. I retaliate by convincing his friends that he is oppressing them and they need to start a communist revolution to kick him out of the country, which works. Later he gets back in and convinces his friends to join my country under fake names, swelling the ranks of voters with people who are there just to vote for the worst policies in order to destroy the country. This becomes so bad that my friend Evan pulls a bloodless coup to abolish democracy and make himself sole leader, but then he cracks down so hard on John’s supporters that everyone gets upset and leaves (“emigrates”). This upsets my friend Bill, who somehow hacks John and tries to delete all his stuff; John counterhacks Bill and destroys his country. Then we all team up with a bunch of guys from Ireland, infiltrate John’s country and destroy it the same way he destroyed us as an act of revenge.

All this happened within about three months real-time, and I was in this hobby for ten years. *Ten years*.

There was an entire era when people would accuse other people of having said racist things on IRC (where logs were often unavailable, and context was absent). This would then be followed with the demand that every political ally of the affected person shun him forever and kick him out of the country and destroy every institution he had built, or else

obviously they were secretly racist themselves. This was met with the only possible response: "actually, no, *you're* the one who said racist things on the chat!". These accusations often resembled the MsScribe story in their sheer not-entirely-social-justice-movement-approved incongruity: "You're racist, and you're a fat lardass!" "Oh yeah? Well you're a f**king homophobic autistic Aspie who will never get laid!" Inevitably the more popular person would win and anyone so foolish as to defend the unpopular person (which I *kept doing*, because I never learn) was banished to Racist Hell. As for Actual Hell, there was a guy named Archbishop Fenton who kept saying really extreme Christian stuff about how we were all going there, and although we all suspected he was a sockpuppet I was never able to figure out whose.

So MsScribe? I'll give her this: she was a gifted amateur. That is it. An amateur. We had frickin' decade-old "intelligence organizations" whose entire job was to collect a network of spies – some real people, some sock puppets – who would join other people's countries under fake (or real!) identities, get information on their secret plans, and throw important elections in favor of the parties we supported. I'm not even ashamed of my role leading one of the largest of these organizations, Shireroth's spy bureau S.H.I.N.E. – if we had unilaterally disengaged from these kinds of games, we would have been demolished by people who didn't.

I remember our scandals. We would build up "dossiers" on various individuals, then publicize them at times calculated to cause maximum damage. One of my favorite was when a prominent female politician was revealed to in fact be male – causing her support to plummet among the key "people who do whatever a girl says in the hopes that she will like them" demographic. In another, which happened a bit after my semi-

retirement, the micronational world's largest communist country, with thirty highly active citizens and a prominent international role, was found to be just one guy posting under thirty different names.

As leader of an espionage organization, I was expected to be able to avoid these damaging revelations, advise my countrymen on how to do the same, and run circles around my enemies. Without tooting my own horn too much, I maintained my most successful character for the better part of a year. This was a guy named Yvain, who infiltrated a Celtic-themed fantasy state called the Duchy of Goldenmoon, took it over, took over its largest neighbor, and was halfway to ultimate power over the entire continent before I got accepted to medical school and decided I should probably reassess how I was using my time.

(to create a paper trail and avoid breaking character, I used the nick "Yvain" for a lot of the websites I joined around this period, which is why half the Internet *still* knows me by that name. I am suitably embarassed by this)

Now I will say this for us boys – and we were boys, like 95% of us, and even the girls were usually found to be boys after careful investigation. We did it with class, we did it with cool names like "Paramountgate" and "The Three Hours' War", we wrote up our petty scandals into epic history books with bibliographies and appendices, and we backstabbed each other so elegantly it would make Machiavelli shed a single tear of pure joy. But in the end? We behaved *exactly* like teenage girls in a Harry Potter fandom.

It is hard at this point not to be reminded of the Robbers' Cave experiment. Social psychologists divided boys at a camp into two groups, intending to do some experiments in order to

figure out what they needed to do to make the groups hate each other, only to learn that the boys had *already* started hating each other with the burning fire of a thousand suns while they were busy planning the experiments. They boys had even formed little group identities, like "Our group are the rough and street-smart ones, the other group is a bunch of holier-than-thou goody-goodies" (the groups were chosen at random).

I read a lot of psychology even as a teenager, so it never surprised me that separating people out into different fictional countries would have the same effect.

But it did kind of surprise me that you could get *quite* those depths of hatred between people who thought that a fictional wizard should hook up with his best friend, versus other people who who thought he should hook up with his other best friend's little sister. Every time I feel like my opinion of people is sufficiently low, I get new evidence making me bump it lower.

Anyway, once those depths of hatred are established, they will proceed in the same way among twenty-somethings trying to discuss Harry Potter romantic pairings, teenagers trying to run fictional countries, and Senators trying to pass vitally important legislation. And that's why, if aliens ever requested exactly one item to teach them about the human race, I would give them the MsScribe story.

They'd kill us all, of course. They would sterilize Earth so thoroughly that not even the archaeobacteria would remain. But in the moment before I was vaporized, I would feel like our species had finally been *understood*.

The Spirit of the First Amendment

Popehat comments on some of the same issues I brought up yesterday from the opposite point of view. They bring up an interesting idea they call the "Doctrine Of The Preferred First Speaker":

> The phrase "the spirit of the First Amendment" often signals approaching nonsense. So, regrettably, does the phrase "free speech" when uncoupled from constitutional free speech principles. These terms often smuggle unprincipled and internally inconsistent concepts — like the doctrine of the Preferred First Speaker. The doctrine of the Preferred First Speaker holds that when Person A speaks, listeners B, C, and D should refrain from their full range of constitutionally protected expression to preserve the ability of Person A to speak without fear of non-governmental consequences that Person A doesn't like. The doctrine of the Preferred First Speaker applies different levels of scrutiny and judgment to the first person who speaks and the second person who reacts to them; it asks "why was it necessary for you to say that" or "what was your motive in saying that" or "did you consider how that would impact someone" to the second person and not the first. It's ultimately incoherent as a theory of freedom of expression.

In other words, person A is within their Constitutional rights to rant about how much they hate gays, person B is within their Constitutional rights to go on a rant about how much they hate person A, and if you condemn person B's speech as "hateful"

or “unnecessary” you’re ignoring the basic symmetry of the situation.

It’s a very well-framed idea, and I remember trying to grope towards something like it when I read Michael Anissimov’s Jezebel’s Vigilante Squad on *More Right*. Parts of that post struck me the wrong way, as genuine “Doctrine of Preferred First Speaker” examples, and I have no doubt that Popehat is complaining about a real thing that some people do.

But in the end I have to disagree with Popehat. I think there *is* a legitimate meaning to “spirit of the First Amendment”, I think it rescues parts of Michael’s post on Jezebel, I think it rescues some of the people defending Phil Robertson, and I think it ends up being a really important part of free speech in general.

What is the “spirit of the First Amendment”? Let’s ask Less Wrong:

> There are a very few injunctions in the human art of rationality that have no ifs, ands, buts, or escape clauses. This is one of them. Bad argument gets counterargument. Does not get bullet. Never. Never ever never for ever.

Why is this a rationality injunction instead of a legal injunction? Because the point is protecting “the marketplace of ideas” where arguments succeed based on the evidence supporting or opposing them and not based on the relative firepower of their proponents and detractors. And as I mentioned yesterday, we’re not talking some theoretical ivory tower idea here, we’re talking about things like how support for gay marriage has increased by an order of magnitude over the past few decades.

What does “bullet” mean in the quote above? Are other projectiles covered? Arrows? Boulders launched from catapults? What about melee weapons like swords or maces? Where exactly do we draw the line for “inappropriate responses to an argument”?

A good response to an argument is one that addresses an idea; a bad argument is one that silences it. If you try to address an idea, your success depends on how good the idea is; if you try to silence it, your success depends on how powerful you are and how many pitchforks and torches you can provide on short notice.

Shooting bullets is a good way to silence an idea without addressing it. So is firing stones from catapults, or slicing people open with swords, or gathering a pitchfork-wielding mob.

But trying to get someone fired for holding an idea is *also* a way of silencing an idea without addressing it. I’m sick of talking about Phil Robertson, so let’s talk about the Alabama woman who was fired for having a Kerry-Edwards bumper sticker on her car (her boss supported Bush). Could be an easy way to quiet support for a candidate you don’t like. Oh, there are more Bush voters than Kerry voters in this county? Let’s bombard her workplace with letters until they fire her! Now she’s broke and has to sit at home trying to scrape money together to afford food and ruing the day she ever dared to challenge our prejudices! And the next person to disagree with the rest of us will think twice before opening their mouth!

The e-version of this practice is “doxxing”, where you hunt down an online commenter’s personally identifiable information including address. Then you either harass people they know personally, spam their place of employment with

angry comments, or post it on the Internet for everyone to see, probably with a message like "I would never threaten this person at their home address *myself*, but if one of my followers wants to, I *guess* I can't stop them." This was the Jezebel strategy that Michael was most complaining about. Freethought Blogs is also particularly famous for this tactic and often devolves into sagas that would make MsScribe herself proud.

A lot of people would argue that doxxing holds people "accountable" for what they say online. But like most methods of silencing speech, its ability to punish people for saying the wrong things is entirely uncorrelated with whether the thing they said is actually wrong. It distributes power based on who controls the largest mob (hint: popular people) and who has the resources, job security, and physical security necessary to outlast a personal attack (hint: rich people). If you try to hold the Koch Brothers "accountable" for muddying the climate change waters, they will laugh in your face. If you try to hold closeted gay people "accountable" for promoting gay rights, it will be very easy and you will successfully ruin their lives. Do you really want to promote a policy that works this way?

There are even more subtle ways of silencing an idea than trying to get its proponents fired or real-life harassed. For example, you can always just harass them online. The stronger forms of this, like death threats and rape threats, are of course illegal. But that still leaves many opportunities for constant verbal abuse, crude sexual jokes, insults aimed at family members, and dozens of emails written in all capital letters about what sorts of colorful punishments you and the people close to you deserve.

Right about the time I started investigating the atheist blogosphere, one popular atheist blogger – I can't remember

her name, but I think she was also on Freethought Blogs – shut down her blog after getting an unmanageable number of these. Everyone posting these messages was entirely within their constitutionally protected right to free speech, yet *something went wrong*. A strong voice for atheism was silenced not because her opponents had clever ideas that contradicted her points, but because they managed to harass her off the podium.

Sometimes this can happen by accident – a no-name nobody makes a statement, a very popular blog or the media picks up on it and broadcasts it, and suddenly thousands of people descend on that person telling them how wrong they are. This is nobody's fault – each individual is completely within their rights to counterargue – but in aggregate it is equivalent to the worst harassment you have ever undergone times ten, and you're always afraid one of those thousands of people is going to take it upon themselves to contact your employer or your family or something. I don't have a good solution to this other than to mention that it is a supererogatory but important duty not to join in.

My answer to the "Doctrine Of The Preferred First Speaker" ought to be clear by now. The conflict isn't always just between first speaker and second speaker, it can also be between someone who's trying to debate versus someone who's trying to silence. Telling a bounty hunter on the phone "I'll pay you $10 million to kill Bob" is a form of speech, but its goal is to silence rather than to counterargue. So is commenting "YOU ARE A SLUT AND I HOPE YOUR FAMILY DIES" on a blog. And so is orchestrating a letter-writing campaign demanding a business fire someone who vocally supports John Kerry.

Bad argument gets counterargument. Does not get bullet. Does not get doxxing. Does not get harassment. Does not get fired

from job. Gets counterargument. Should not be hard.

A Response to Apophemi on Triggers

[content warning: discussion of triggers. Mentions various triggers. Mentions, without using or condoning, racial slurs]

I.

I originally planned not to respond to Apophemi's essay requesting that people not discuss potentially triggering ideas dispassionately, because my response would inevitably have to discuss a lot of triggering ideas, and it would be dispassionate, and that *might* not be the most effective way of conveying that I take zir concerns seriously.

(Apophemi's essay complains about being misgendered but doesn't give me ironclad evidence what zir gender is, so I'm going to use the gender neutral pronoun here as a least bad option. No offense is intended and if Apophemi tells me what pronoun ze prefers I'll edit it in.)

I'm changing my mind for two reasons. First, everyone else is doing it, so Apophemi has probably reached Peak Triggering by now and the situation can't get any worse. Second, I feel like it would be more respectful and productive to object and give zir a chance to explain why my objections are wrong, than to just say "I disagree with this but I'm not going to explain why" and dismiss the whole thing outright.

(That having been said, if Apophemi doesn't want to read this, I am totally in favor of this; ignoring all posts on my blog tagged "race/gender/etc" is always a good life choice.)

My one worry is the comment thread. I no longer trust my commenters to be kind or reasonable, and since we're talking specifically about triggers and giving a big list of triggery things, unkind people present a problem. So I am closing

comments for this thread. If Apophemi wants to make a response, ze may email it to me or post it somewhere and I will add it in.

II.

Apophemi writes:

> On the other hand, most of these things involve warning signs for opinions whose holders are frequently detrimental to my health and safety, and therefore I feel pretty entitled to these boundaries, and pretty insulted at the implication that possessing such boundaries is inferior to not possessing them.
>
> An example: I cannot in good faith entertain the argument that high-scarcity societies are right in having restrictive, assigned-sex-based gender roles, even if these social structures result in measurable maximized utility (i.e. many much kids). I have a moral imperative against this that overrides my general impulse towards maximized utility, or rather (if you asked me about it personally) tilt-shifts my view of what sectors 'deserve' to see their utility maximized at the expense of a given other sector.
>
> However, this results in a knee-jerk intellectual squick when I run across someone entertaining or endorsing these arguments. (If I were being YouTube-commenter-style punchy about this, this entire post would have been a comment reading "'Fertile women' my ass.", for the record.) This is because respect for said arguments and/or the idea behind them is a warning sign for either 1) passively not respecting my personhood or 2) actively disregarding my personhood, both of which are, to use some vernacular, hella fucking dangerous to me personally.

> I am reasonably confident (insert p value here) that this attitude is self-replicating among people who are accustomed to being at risk in a specific way that generally occurs to marginalized populations. (I cannot speak for people who may have a similar rhetorical roadblock without it being yoked to a line of social marginalization, other than that I suspect they happen.) This would mean that rewarding the "ability" to entertain any argument "no matter how 'politically incorrect'" (to break out of some jargon, "no matter how likely to hurt people") results in a system that prizes people who have not been socially marginalized or who have been socially marginalized less than a given other person in the discussion, since they will have (in general) less inbuilt safeguards limiting the topics they can discuss comfortably.
>
> In other words, prizing discourse without limitations (I tried to find a convenient analogy for said limitations and failed. Fenders? Safety belts?) will result in an environment in which people are more comfortable speaking the more social privilege they hold. (If you prefer to not have any truck with the word 'privilege', substitute 'the less likelihood of having to anticipate culturally-permissible threats to their personhood they have lived with', since that's the specific manifestation of privilege I mean. Sadly, that is a long and unwieldy phrase.)

This reminds me of the idea of safe spaces.

Safe spaces are places where members of disadvantaged groups can go, usually protected against people in other groups who tend to trigger them, and discuss things relevant to

that group free from ridicule or attack. I know there are many for women, some for gays, and I recently heard of a college opening one up for atheists. They seem like good ideas.

I interpret Apophemi's proposal to say that the rationalist community should endeavor to be a safe space for women, minorities, and other disadvantaged groups.

One important feature of safe spaces is that they can't always be safe for two groups at the same time. Jews are a discriminated-against minority who need a safe space. Muslims are a discriminated-against minority who need a safe space. But the safe space for Jews should be very far way from the safe space for Muslims, or else neither space is safe for anybody.

The rationalist community is a safe space for people who obsessively focus on reason and argument even when it is socially unacceptable to do so.

I don't think it's unfair to say that these people need a safe space. I can't even count the number of times I've been called "a nerd" or "a dork" or "autistic" for saying something rational is too high to count. Just recently commenters on Marginal Revolution – not exactly known for being a haunt for intellect-hating jocks – found an old post of mine and called me among many other things "aspie", "a pansy", "retarded", and an "omega" (a PUA term for a man who's so socially inept he will never date anyone).

The reason the rationalist community tends to talk about controversial issues like race and gender on occasion is that the whole point of rationalism is giving things a fair analysis regardless of whether it's socially popular or acceptable to talk about. So of *course* it will start focusing on all of the ideas that are least acceptable to talk about. I remember talking to

someone who admitted, after several false starts and awkward pauses, that he found the scientific research on differences between races pretty convincing. I answered that I was still neutral on the matter but that Jensen was indeed a pretty darned meticulous researcher, and he very nearly cried with relief. He'd thought he was a terrible person for taking the research seriously, had never been able to talk about it with anyone, was stuck in a guilt spiral over it, and I was the first person to give him basic human sympathy.

And I think most people in the rationalist community have shared this reaction – not necessarily about race and gender issues, because contrary to the above we really don't talk about those that much – but about atheism, or transhumanism, or negative utilitarianism, or simulationism, and they had finally found people who would pay them the respect to debate their ideas on merit instead of mouthing the appropriate social platitude to dismiss it as horrible or as totally obvious.

If you are the sort of person with the relevant mental quirk, living in a society of people who don't do this is a terrifying an alienating experience. Finding people who are like you is an amazing, liberating experience. It is, in every sense of the word, a safe space.

If you want a community that is respectful to the triggers of people who don't want to talk about controversial ideas, the Internet is full of them. Although I know it's not true, sometimes it seems to me that half the Internet is made up of social justice people talking about how little they will tolerate people who are not entirely on board with social justice ideas and norms. Certainly this has been my impression of Tumblr, and of many (very good) blogs I read (Alas, A Blog comes to mind, proving that my brain sorts in alphabetical order). There

is no shortage of very high-IQ communities that will fulfill your needs.

But you say you're interested in and attracted to the rationalist community, that it would provide something these other communities don't. Maybe you are one of those people with that weird mental quirk of caring more about truth and evidence than about things it is socially acceptable to care about, and you feel like the rationalist community would be a good fit for that part of you. If so, we would love to have you!

But if you want to join communities specifically because they are based around dispassionate debate and ignoring social consequences, but your condition for joining is that they stop having dispassionate debate and take social consequences into account, well, then you're one of those people – like Groucho Marx – who refuses to belong to any club that would accept you as a member.

Imagine a Jew walking into a safe space for Muslims, and saying he finds Islam really interesting and wants to participate, but that in order for it to be a safe space for him they really need to stop talking about that whole "Allah" thing.

III.

I deliberately said "the rationalist community" above rather than "Less Wrong", because Less Wrong explicitly *does* try to be a safe space. It has a (vague and very poorly enforced) ban on talking about politics or other controversial topics which successfully discourages Reactionaries and their ilk from starting threads directly about their controversial views (they often get away with discussing other results that refer to them only indirectly).

These topics nevertheless come up anyway at regular intervals. There is almost always the same pattern when this happens:

A feminist or other person in the social justice movement very prominently posts a declaration that everyone on the site needs to be more feminist and social-justice-y. They get heavily upvoted.

A few people in the comments politely disagree, sometimes with the gist of the post, other times with specific claims.

Other people express outrage that anyone would disagree, and say this just proves that the site is full of horrible people and that feminism and social justice are needed now more than ever.

World War III happens.

It happened when Daenerys gathered a whole series of feminist things from people that got posted to Discussion called things like "On Creepiness" and "On Misogyny". It happened when Multiheaded, a Marxist somewhere to the left of Kropotkin, posted a thread complaining about people complaining that there were people complaining about controversial opinions on the site (or something). It happened when Apophemi's essay itself got posted to the site and heavily upvoted.

I've downvoted all of these things, not because I disagree with them (although I often do) but because the ban on politics is really useful to avoid exactly this kind of situation. I hope in the future it is more consistently enforced, and I hope this would be more conducive to the kind of site Apophemi wants.

"But," people object "banning politics is hard, and talking politics sometimes is fun, and besides, social justice ideas are important to disseminate. Can't we just ban the nasty, triggering kinds of politics?"

This would be a good time to admit that I am massively, *massively* triggered by social justice.

I know exactly why this started. There was an incident in college when I was editing my college newspaper, I tried to include a piece of anti-racist humor, and it got misinterpreted as a piece of pro-racist humor. The college's various social-justice-related-clubs decided to make an example out of me. I handled it poorly ("BUT GUYS! THE EVIDENCE DOESN'T SUPPORT WHAT YOU'RE DOING!") and as a result spent a couple of weeks having everyone in the college hold rallies against me followed by equally horrifying counter-rallies for me. I received a couple of death threats, a few people tried to have me expelled, and then everyone got bored and found some other target who was even more fun to harass. Meantime, I was seriously considering suicide.

But it wasn't just that one incident. Ever since, I have been sensitive to how much a lot of social justice argumentation resembles exactly the bullying I want a safe space from – the "aspie", the "nerd", that kind of thing. Just when I thought I had reached an age where it was no longer cool to call people "nerds", someone had the bright idea of calling them "nerdy white guys" instead, and so transforming themselves from schoolyard bully to brave social justice crusader. This was the criticism I remember most from my massive Consequentialism FAQ – he's a nerdy white dude – and it's one I have come to expect any time I do anything more intellectual than watch American Idol, and usually from a social justicer.

(one reason I like the MsScribe story so much is that it really brings into relief how aligned social justice and bullying can be. I'm not saying that all or even most social justice is about bullying. Just *enough*)

The worst part was when I read some social justice essay – I can't remember where – which claimed that it was impossible to bully a member of a privileged group. That it didn't count. That there was no such thing. So not only did they *sound* suspiciously like bullies, but they were conveniently changing the rules so that it was impossible by definition for me to be bullied at all, and all my friends (except for the black ones) who had problems with bullies as a child or in the present – didn't count, didn't exist, didn't deserve any sympathy.

I believe you mentioned in your essay that feeling like you're being told you're not a person is really scary? Well, just so.

So suffice it to say I am triggered by social justice. Any mildly confrontational piece of feminist or social justice rhetoric sends me into a panic spiral. When I read the essay this post was based on, I got only about four hours of sleep that night because my mind was racing, trying to figure out whether I was going to get in trouble about it and whether anyone who supported it could hurt me and how I could defend myself against it.

Because my mind doesn't just let me feel sad for a minute and then move on – no, that would be too easy. It gives me this massive compulsion to "defend myself" against any piece of social justice I see by writing really long and complete rebuttals. Which inevitably attract more social justice people wanting to debate me. And unfortunately, outrage addiction is a very real thing, and I find myself *actively seeking out* the most horrible social justice memes in order to be horrified by them.

(…also, telling me I'm not allowed to be triggered by my triggers is itself a trigger. Whoever designed the human mind was *really* kind of a jerk.)

I struggle against this all the time. H.L. Mencken writes "Every normal man must be tempted at times to spit on his hands, hoist the black flag, and begin to slit throats." Well, this is my temptation. It requires more willpower than anything else I do in my life – more willpower than it takes for me to get up in the morning and work a ten hour day – to resist the urge to just hoist the black flag and turn into a much less tolerant and compassionate version of Heartiste.

I don't think I'm at all alone in this. Like, you may notice there's a large contingent of people – mostly men, but a surprising number of women as well – who totally freak out when they hear social justice stuff and seem to loathe social justice with an unholy passion? And maybe you've wondered whether the classic glib dismissal of them as people benefitting from the patriarchy who are upset about "uppity women" *quite* explains the level of rage and terror and sudden lashing out?

If you are, indeed, someone who has been traumatized and is easily triggered, you can probably recognize the signs yourself. There's a certain desperation, a certain terror thinly disguised by rage that doesn't really come from anything else.

So suffice it to say I am triggered by social justice, and probably a lot of other people are too. Why do I make such a big deal of this?

First, because it has a lot of bearing on whether we can just ban triggery things. There is a certain school of thought that there are two or three excessively evil things that trigger other people, like making fun of rape, and once we make people stop those, we will live in a trigger-free paradise.

But that's not true. I'm triggered by feminism. My girlfriend is also triggered by certain kinds of feminism (long story), but

also by many discussions of charity – whenever ze hears about it, ze starts worrying ze is a bad person for not donating more money to charity, has a mental breakdown, and usually ends up shaking and crying for a little while.

Since we can never make every form of discussion respect everybody's triggers, that leaves two solutions. First, we can try the "my triggers are important, your triggers are invalid" solutions and end up with powerful groups able to enforce their triggers, and weak groups being told to "just man up". Second, we can try the safe space solution, where not everyone can be certain of safety everywhere, but everyone is certain of safety somewhere. I don't expect Tumblr to stop being feminist for me, but I *have* managed to scrub my Facebook feed so thoroughly that I only get about two or three articles per week on how hilarious it would be if male superheroes were dressed like female superheroes. One learns to relish little victories.

Second, because I think the essay contains a false dichotomy: *privileged* people don't have any triggers, *oppressed* people do. *You guys* are intact, *I* am broken. But truth is, everybody's broken. The last crown prince of Nepal was raised with limitless wealth and absolute power, and he still freaked out and murdered his entire family and then killed himself. There's probably someone somewhere who still believes in perfectly intact people, but I bet they're not a psychiatrist.

Third, because I have not yet raised the black flag. And some of my resistance I credit to – the rationalist community. The Litany of Tarski: "If all feminism is horrible, I desire to believe that all feminism is horrible. If all feminism is *not* horrible, I desire to believe that all feminism is not horrible." It is a calming litany. Sometimes it helps.

A Christian proverb says: “The Church is not a country club for saints, but a hospital for sinners”. Likewise, the rationalist community is not an ivory tower for people with no biases or strong emotional reactions, it’s a dojo for people learning to resist them.

I do not think it is always wrong for people to engage in activities that exclude certain categories of disadvantaged people. For example, music naturally excludes the deaf (someone will bring up Beethoven here. You know what I mean). Horseback riding excludes most people too poor to buy horses or live in the country. This is sad, but these activities should still continue.

But I do not think dispassionate discussion for the easily triggered is as bad as all that. It is more like marathons for people who are out of shape. They will have difficulty at first. If they want to learn, they can. If they try, they will become stronger. I can’t run a marathon and I can’t always discuss issues fairly and dispassionately, but I’m glad both activities exist as things to aspire to.

IV.

> Following sufficient rinsing and repetition, it may occur to someone in a ‘discourse without limitations’ community to wonder where all the (say) queer people and/or women and/or trans* people and/or disabled people and/or people of color and/or non-American-and-Northern-European people and/or citizens of the third world and/or people whose first language is not English and/or Jewish people and/or etc. (repeat and/or for any population ‘coincidentally’ discouraged from participating) went.

> Or rhetorical-you could argue that women and/or minorities and/or historically disadvantaged groups are inherently irrational / otherwise not qualified for community membership, at which point I would proceed to avoid rhetorical-you, as above.

You are implying – not saying, but I hope it is fair to read the implication – that "discourse without limitations" drives minority group members away from the communities that participate in it.

This has recently become an interest of mine because a number of communities I'm in – the atheist community, the rationalist community, the Reddit community, the Vaguely Techy Bay Area community – notably lack certain kinds of minorities. And there are many people who say this must be because of some kind of inherent flaw in the community, that it proves either that community members are racist, or at least that they are less actively non-racist than might be desired. Sometimes people say this nicely and helpfully, like you have. Other times people say it more confrontationally, often with the standard "nerdy white dudes" line thrown in.

And always they make the same dichotomy you do – between the "driving these people away" explanation, and the "are you claiming these people are inherently inferior?" explanation. And the proposed solution is always to be more "respectful", which means talking more about feminism and social justice, and being less accepting of people who counterargue against it.

Needless to say, this is not a solution I can entirely get behind. So I am terribly biased on this point. Still, let me nevertheless present my argument for evaluation.

I have been to several yoga classes. The last one I attended consisted of about thirty women, plus me (this was in Ireland; I don't know if American yoga has a different gender balance).

We propose two different explanations for this obviously significant result.

First, these yoga classes are somehow driving men away. Maybe they say mean things about men (maybe without intending it! we're not saying they're intentionally misandrist!) or they talk about issues in a way exclusionary to male viewpoints. The yoga class should invite some men's rights activists in to lecture the participants on what they can do to make men feel comfortable, and maybe spend some of every class discussing issues that matter deeply to men, like Truckasaurus.

Second, men just don't like yoga as much as women. One could propose a probably hilarious evolutionary genetic explanation for this (how about women being gatherers in the ancestral environment, so they needed lots of flexibility so they could bend down and pick small plants?) but much more likely is just that men and women are socialized differently in a bunch of subtle ways and the interests and values they end up with are more pro-yoga in women and more anti-yoga in men. In this case a yoga class might still benefit by making it super-clear that men are welcome and removing a couple of things that might make men uncomfortable, but short of completely re-ordering society there's not much they can do to get equal gender balance and it shouldn't be held against them that they don't.

The second explanation seems much more plausible for my yoga class, and honestly it seems much more plausible for the rationalist community as well.

We're not actually missing all those groups you mention as minorities who might be driven away. In fact, in many cases, we have *far* more of them than would be expected by chance. For example, we contain transgender people at about five times the rate in the general population (1.5% vs. 0.3%), and gays/lesbians/bisexuals at about three times the rate in the general population (15% vs. 4%). People who Jewish by descent are four times the national average (8% vs. 2%), and people with mental disorders are either around equal to the general population or much much higher, depending on how one interprets the data I did a terrible job collecting (sorry). We have more people with English as a second language than almost any other online community I know (the country with most rationalists per capita continues to be Finland) and members from Kenya, Pakistan, Egypt, and Indonesia.

The only groups we appear to be actually short on are women and minorities (and then only if you follow standard American practice of refusing to count Asians as a real minority, numbers be damned).

But just as you would not immediately jump from the overrepresentation of transsexuals to the assumption that we must somehow be discriminating against cissexual people, so one does not jump from the overrepresentation of men to the assumption that women are being discriminated against.

Most rationalists come from the computer science community, which is something like 80% male. A few come from hard science fields like math and physics, both of which are 80 – 90% male. There is *zero* need to invoke "discourse without limitations" as an explanation for why the rationalist community is heavily male-dominated, and any attempt to do so would run into the question of why the occasional dispassionate cost-benefit discussion of eugenics apparently

horrifies women but heavily attracts Jews, gays, and people with mental disorders.

Worse, the hypothesis fails in the other direction as well. There are lots of groups that are horribly offensive towards minorites yet nevertheless manage to have very many of them. Across nearly every denomination, far more women than men go to church – if you go to a Catholic Mass, you will see pews full of ladies at levels the atheist community can only dream of. The atheist community is so feminist that there has been a serious movement to replace it with "Atheism Plus" that excludes all non-feminists; the Catholic Church is so regressive that it won't let women become priests and thinks they were created as a "helpmeet" for man. And yet women, in aggregate, love the one and hate the other.

You know what other community has more women than the rationalist community? The men's rights movement. According to the /r/mensrights survey, about 9.3% of men's rights activists are female, which is slightly fewer women than the rationalist community on the last survey, but slightly more women than the rationalist community on the survey before that. A friend who reads Heartiste guesses that about a third of his commenters are female (though adds that some of these may be men who are pretending in order to make a point). So if we actually spent all our time belittling women and justifying their oppression, as far as I can tell our percent female readership would probably go *up*.

I am left pretty certain that the male-dominated rationalist community has a gender imbalance for the same reason as my female-dominated yoga class. We could always see whether it might help to inviting some feminists in, listen to them without protest, and agree to do whatever they say – but I would enjoy that about as much as you would enjoy getting lectured by

men's rights activists without being able to protest, and the end result would probably be about the same.

V.

Apophemi's essay continues with an addendum:

> there's significant linguistic signalling that can make up the difference between people who have more to lose from apparently innocent argument participating or not. For (specific to my experience) example…
>
> – arguments against accusations of racism/sexism/cissexism/heterocentrism/ableism/etc. that boil down to "those are silly words and they aren't in my spellcheck"

I worry that you're not being entirely fair here. Who the heck doesn't have "racism" in their spellcheck? I feel like your opponents may be making a more subtle point than you think.

When I ask people to use words other than "racism", it's usually because I believe a Worst Argument In The World is being sprung on me – the article will explain more. I think this is a reasonable concern, and it's always fair to ask someone to taboo their words.

But there's another problem I sometimes run into with some other concepts, like "male privilege" or "male gaze" or "marginalized".

You said you enjoyed my Anti-Reactionary FAQ (thanks!), so I wonder whether you enjoyed section 2.3.1, in which I deconstruct the word "demotist".

The Reactionaries argued that "demotist" countries, meaning countries that had some notion of popular sovereignty including communisms, non-monarchical dictatorships, and

democracies – had a terrible human rights record. Which is true – Maoist China (communist), Myanmar under the junta (non-monarchical dictatorship) and many others do have terrible human rights records.

But the Reactionaries were loading the debate by using the word "demotist", which deliberately groups those regimes together with stable liberal democracies (who have fantastic human rights records compared to anyone else). My argument here was *exactly* that "demotist" wasn't in my spellcheck and that in order to "win" the debate the Reactionaries had to invent new words that loaded the argument in their favor. Denied the ability to use their own words and forced to use the same vocabulary as the rest of us, their argument totally falls apart.

Not everything must be stated in ordinary language – if you didn't let chemists use terms like "valence electron" or "ionic", you would be denying them a useful tool that makes chemistry much easier. I get that.

But when people are trying to talk about ordinary processes, and they insist on using their own words which don't exactly correspond to features of the world, and they can't always make the same arguments with more standard words, I get *super* suspicious.

Words are hidden inferences. They encode assumptions, and sometimes those assumptions are correct and other times they are wrong. This is true more than usual with jargon, and even more than usual with partisan political jargon (don't call them "rich", call them "job-creators"!) It is useful and acceptable to ask people to take a step back from their words to examine whether the assumptions behind them are correct.

– conflating terms describing marginalization (such as the above) with insults (i.e. “calling me racist is an insult”, “let’s discuss this without using meaningless insults like ‘misogynist’”),
– use of insults that have a history of being specific to women or that effectively mean “this person is like a woman”

Oh God oh god oh god oh god you *so* do not understand oh god.

Which words are or are not slurs is not a feature of the word’s etymology or even the intent of people using those words. For example, the word “Jap”, on its own, is very clearly just a convenient shortening of “Japanese person” in the same way that “Brit” is very clearly a convenient shortening of “British person”.

Yet it is *not* okay to go around calling Japanese people “Japs” and then lecturing them because they are “conflating” a term describing their heritage with an insult (“ha ha, that silly Japanese person thinks I’m insulting her just because I used a shortened form of her demonym”).

Most Japanese people have a history – maybe personal, maybe just second-hand – of *correctly* associating the word “Jap” with an attempt to dehumanize them, marginalize them, or cause them huge amounts of personal grief. It doesn’t matter whether *you* think “Jap” was meant to be offensive, if a Japanese person tells you they’re triggered by it and you keep using it, you’re a jerk.

(and the same is true of a much more famous slur which is a derivation of the perfectly innocent Latin word *niger* meaning “black-colored”, but which has been wrenched *far* away from that perfect innocence by the referents of the term having *more*

than enough opportunities to associate that word with an attempt to hurt them.)

So a neutral word can become an insult or trigger or slur if it is associated sufficiently strongly and sufficiently often with people trying to hurt you.

Now, when those people were sending me death threats because of that article in the college paper, what word do you think they used?

When the media talks about a "scandal" in which some politician or actor is accused of being offensive and then gets fired from their job and has to do a live apology on national TV during which they break down crying, what word do you think always starts the process?

When you read the MsScribe story – which a dozen people in the comments said struck incredibly true to life for them – what word did MsScribe use to deride her enemies before kicking them out of the community and making everyone refer to them as "cockroaches" and posting sexually explicit stories about them doing horrible things?

People have an *incredibly reasonable terror* of that specific word, and when you refuse to change it to one of many dozens of available synonyms, that has some pretty strong implications about where you are coming from. It says "I don't respect you enough not to use this word that terrifies and triggers you" which in turn means that people's terror and triggering is probably correct.

I am sure there are some lovely elderly Southerners who use [the n-word] simply because that was what they grew up with, and are mildly annoyed every time a black person throws a fuss about it because they honestly didn't mean any harm. And they use that exact same argument: "I didn't mean anything by

it, it's just what I call people like you, you're so sensitive treating it as an insult." But they are missing the point. It doesn't matter what *their* feelings are, it matters whether it hurts other people.

And when they *anticipate* this, like "Oh, I'm going to call that black person the n-word, and I bet he's going to get all upset about it, you know how they are", *that* doesn't seem innocent to me. It sounds like they know they're hurting other people and just don't care.

And when you say you *expect* people to feel insulted and triggered by the word "racist", but you're going to do it anyway, even though you are perfectly aware of other words you could use that would actually be more descriptively accurate, I kind of have the same worry.

And then your very next point is that you don't want people to use terms you consider slurs. Well, yes, of course that is fair! And I try to avoid slurs as much as I can.

Yet I cannot help rounding this entire section off to "The two things that annoy me are when other people use language that triggers me, and when other people ask me to stop using language that triggers *them*."

And when I have bring this up to people, they usually answer "It's impossible to trigger a member of a privileged group" or "Triggering a member of a privileged group doesn't count". I am so happy you have defined away my pain. THIS IS THAT DEHUMANIZATION THING AGAIN.

In conclusion, aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaah.

> Sense of being persecuted by "political correctness"…I have failed so far to find a definition of "political correctness" in this context that could not be search-and-

> replaced with "trying to avoid hurting people" to either no effect or increased comprehensibility. You are free to attempt to change my mind on this, I guess?

I like this sentence because it is a good example of language in fact making a difference, of words being hidden inferences, of reasonable requests not to use terms that don't just boil down to "it isn't in my spellcheck". In fact, Scott Adams makes this exact point in his essay What's The Difference Between A Sexist And A Regular Asshole? It intrigues me that both sides are trying to remove the others' linguistic weapons by demanding they be deactivated and replaced with normal words, but are refusing to relinquish their own. Anyhow…

When I see references to "political correctness", it's usually followed by something like "has gone too far". This suggests a reasonable interpretation to me – political correctness is indeed a way of trying to avoid hurting people, but like all forms of trying to avoid hurting people, it can go too far.

Trying to prevent terrorism is good. But when any vaguely Muslim looking person who tries to board a plane tends to get hauled off and strip-searched without so much as an apology, one can ask whether the legitimate goal of trying to prevent terrorism has gone too far.

Likewise, when people start saying that it's cultural appropriation to eat latkes or a ten-year old girl can be charged with rape for playing a game of Doctor or heterosexual white people can't be depressed or any of the other three million things of this sort I see on Tumblr every day, then I do think it's fair to say that the legitimate goal of trying to protect disadvantaged groups is going too far *in certain cases*.

This is not to say that it has uniformly gone too far in every aspect of society, just that in these cases – the ones the people

saying this have encountered – it has *locally* gone too far.

I do not really know what claim you are asserting – that political correctness never goes too far? That no one trying to protect the rights of minority groups has ever overstepped good sense? This seems more like cheering on a side than stating a defensible position to me.

See also Section II of this essay.

> – pretty much any usage at any point of the word "insane" when we are not talking about a court case by now

Dammit, you just broke my girlfriend.

Ze is a mentally ill person who has attempted suicide a few times and been in and out of mental hospitals, and ze is now *seething* with anger. I think I see smoke coming out of zir ears. Now ze is *demanding* that I write an extremely angry response saying HAVEN'T YOU EVER READ ANYTHING IN THE DISABILITY RIGHTS COMMUNITY??!?! and DON'T YOU KNOW THAT LOTS OF PEOPLE USE TERMS LIKE "INSANE" AND "CRAZY" TO AVOID MEDICALIZING THEIR DISABILITY??!?! and HOW DARE YOU PURPORT TO SPEAK FOR ALL OF US WHO THE HELL APPOINTED YOU OUR REPRESENTATIVE??!

As a psychiatrist myself, avoiding medicalizing disability is not really high on my list of priorities. But as a "mentally ill person" myself, – two years on Paxil followed by eight on Prozac followed by two years of behavioral psychotherapy followed by the *incredibly enjoyable* process of finding a hospital that wanted to employ a psychiatrist with a mental illness because I knew I wouldn't be able to hide it through however many years of work I will be with them – avoiding

the use of the term “insane” has never been high on my list of priorities either.

I have never, ever, noticed the pattern you have – people who use the word “insane” being otherwise bad or de-legitimizing people with mental illness. In fact, it has happened more than once to me – twice, I think, spookily similar – that a mentally ill patient asks me what my diagnosis is, I say something like “schizophrenia” or whatever, and they say “Nope! I’m insane! If you want to be a good doctor, you’re going to have to learn to tell it like it is!”

Both my girlfriend and I agree that people being very concerned about people using or not using specific very common words has been a much bigger warning flag of someone who is otherwise not a nice person than use of the word “insane”.

…which is not to say that you haven’t had the exact opposite experience! That’s kind of the problem. No one can speak for an entire community and community members have very different experiences and preferences. My policy so far has been to always respect someone’s terminology preferences when talking to them personally or in a small group, and to respect terminology preferences I know to be common when talking to a large audience. In a lifetime of working with the mentally ill and dating two different disability rights bloggers I have yet to hear anyone else express a strong preference against “insane”, but if it happens more often I will update. And I will certainly avoid doing it if I ever have reason to talk to you directly.

> If by “sluttiness” r-you mean “sexual promiscuity”, what is gained by using a gender-targeted insult that is likely to make a significant portion (i.e. women and/or queer

people, who together are like… 55% of the world at least) of r-your potential audience uncomfortable and less likely to engage with r-your argument?

This I apologize for unreservedly. In the Anti-Reactionary FAQ, I quoted some reactionary passages using the word "sluttiness", and then I continued using it myself afterwards. I was hoping to kind of mock the reactionaries by pointing out how much their argument depended on that one word. In the process, it seems I offended some women/queers as well. I was wrong to do this, there were very easy ways to avoid it, and I will avoid it in the future.

VI.

You probably aren't even reading this, but I hope someone like you is.

Lies, Damned Lies, and Social Media: False Rape Accusations

[content warning: rape, false rape allegations. Some people have been linking this article claiming it says things it DEFINITELY DOES NOT, so please read it before you have an opinion.]

(see also parts 1, 2, 3, and 4 of ∞)

I.

Spotted on Brute Reason but liked and reblogged 35,000 times: Five Things More Likely To Happen To You Than Being Accused Of Rape. A man is 631 times more likely to become an NFL player than to be falsely accused of rape! Thirty-two times more likely to be struck by lightning! Eleven times more likely to be hit by a comet!

Needless to say, all of these figures are completely wrong, in fact wrong by a factor of over 22,700x. I'm not really complaining – missing the mark by only a little over four orders of magnitude is actually not bad for a "story" of this type. Nevertheless, it will be instructive to figure out where they erred so we may be vigilant against such things in the future, and perhaps certain moral lessons may be gleaned in the process as well.

II.

Since that article itself does not show its work, we will have to rely on its obvious inspiration, an almost-identical blog post written a few days before by the same person responsible for the Buzzfeed piece, Charles Clymer.

It starts by noting that there are about 84,000 forcible rapes per year – and that FBI statistics suggest 8% are false accusations. We will can examine these numbers later, but for now let's just take them as given.

It then goes on to calculate that, given the average man has sex 99 times per year (who *is* this average man?!) there are 5.1 billion acts of sexual intercourse in the United States each year among American men 15 – 39. Divide 5.1 billion by 6,750, and therefore, in Clymer's words "the odds of any sexually-active male between the ages of 15 and 39 has a 750,000 to 1 chance of being falsely accused of rape"

And, he goes on to say, 1/33 men are raped during their lifetime. Therefore, the average man has a 27500x higher chance of being raped than being falsely accused of rape. The average man has a 1 in 84,079 chance of being killed by lightning, so that's 32x more likely than getting falsely accused of rape. And it adds that the average women has a 1/4 chance of being raped during her lifetime – so the odds of a woman being raped during her lifetime must be 220000x higher than the odds of a man being falsely accused of rape.

Did you spot the sleight of hand in those calculations? He calculated the odds of a man who has sex 99 times per year for 24 years being accused of rape *per sex act*, and then declared this was the odds of being accused of rape *in your lifetime*. Then he went on to compare it to various other lifetime odds, like the lifetime odds of being raped, the lifetime odds of being struck by lightning, et cetera.

This isn't comparing apples to oranges. This isn't even comparing apples to orangutans. This is comparing apples to the supermassive black hole at the center of the galaxy.

To highlight exactly how awful this is, suppose we wanted to trivialize rape itself through the same methodology. The average woman, as per the article's statistics, has a 1/4 chance of getting raped during her lifetime, which means a 1/9500 or so chance of getting raped per sex act if she has sex 99 times per year from ages 15-39. And looking at the same list of statistically unlikely things provided on that article, that's less than the odds of dying in a plane crash (1/7032). So you crow "THE AVERAGE WOMAN IS LESS LIKELY TO GET RAPED THAN TO DIE IN A PLANE CRASH! HA HA WOMEN ARE SO DUMB TO EVER WORRY ABOUT RAPE!". And now you have a Buzzfeed article.

III.

We can do better. Let's come up with conservative and liberal estimates of a man's chance of getting falsely accused of rape between ages 15 and 39.

The rate of false rape accusations is notoriously difficult to study, since researchers have no failsafe way of figuring out whether a given accusation is true or not. The leading scholar in the area, David Lisak, explains that the generally accepted methodology is to count a rape accusation as false "if there is a clear and credible admission [of falsehood] from the complainant, or strong evidential grounds", and goes on to explain what these grounds might be:

> For example, if key elements of a victim's account were internally inconsistent and directly contradicted by multiple witnesses and if the victim then altered those key elements of his or her account, investigators might conclude that the report was false

Attempts to use this methodology return varying results. Lisak lists seven studies he considers credible, which find false accusation rates of 2.1%, 2.5%, 3.0%, 5.9%, 6.8%, 8.3%, 10.3%, 10.9%. The two with 10%+ mysteriously go missing and thus we get the commonly quoted number of "two to eight percent", which is repeated by sources as diverse as Alas, A Blog, Slate, and Wikipedia (Straight Statistics keeps the original 2% – 10% number)

Feminists make one true and important critique of these numbers – sometimes real victims, in the depths of stress we can't even imagine, do strange things and get their story hopelessly garbled. Or they suddenly lose their nerve and don't want to continue the legal process and tell the police they were making it up in order to drop the case as quickly as possible. All of these would go down as "false allegations" under the "victim has to admit she was lying or contradict herself" criteria. No doubt this does happen.

But the opposite critique seems much stronger: that some false accusers manage tell their story without contradicting themselves, and without changing their mind and admit they were lying. We're not talking about making it all the way through a trial – the majority of reported rapes get quietly dropped by the police for one reason or another and never make it that far. Although keeping your story halfway straight is probably harder than it sounds sitting in an armchair without any cops grilling me, it seems very easy to imagine that *most* false accusers manage this task, especially since they may worry that admitting their duplicity will lead to some punishment.

The research community defines false accusations as those that can be proven false beyond a reasonable doubt, and all

others as true. Yet many – maybe most – false accusations are not provably false and so will not be included.

So there's reason to believe some of those 2-10% of presumed false accusations are actually true, and other reasons to believe that some of the 98% – 92% of presumed true accusations are actually false.

What is an upper bound on the number of false rape accusations? Researchers tend to find that police estimate 20%-40% of the rape accusations they get to be "unfounded", (for example Philadelphia Police 1968, Chambers and Millar 1983, Grace et al 1992, Jordan 2004, Gregory and Lees 1996, etc, etc). Many scholars critique the police's judgment, suggesting many police officers automatically dismiss anyone who doesn't fit their profile of a "typical rape victim". A police-based study that took pains to avoid this failure mode by investigating all cases very aggressively (Kanin 1994) was criticized for what I think are ideological reasons – they primarily seemed to amount to the worry that the aggressive investigations stigmatized rape victims, which would make them so flustered that they would falsely recant. Certainly possible. On the other hand, if you dismiss studies for not investigating thoroughly enough *and* for investigating thoroughly, there will never be any study you can't dismiss. So while not necessarily endorsing Kanin and the similar studies in this range, I think they make a useful "not provably true" upper bound to contrast with the "near-provably false" lower bound of 2%-10%.

IV.

But this only represents the number of false rape accusations that get reported to the police. 80% of rapes never make it to the police. Might false rape accusations be similar?

Suppose you are a woman who wants to destroy a guy's reputation for some reason. Do you go to the police station, open up a legal case, get yourself tested with an invasive rape kit, hire an attorney, put yourself through a trial which may take years and involve your reputation being dragged through the mud, accept that you probably won't get a conviction anyway given that you have no evidence – and take the risk of jail time if you're caught lying?

Or do you walk to the other side of the quad and bring it up to your school administrator, who has just declared to the national news that she thinks all men accused of rape should be automatically expelled from the college, without any investigation, regardless of whether there is any evidence?

Or if even the school administrator isn't guilty-until-proven-innocent enough for you, why not just go to a bunch of your friends, tell them your ex-boyfriend raped you, and trust them to spread the accusation all over your community? Then it doesn't even *matter* whether anyone believes you or not, the rumor is still out there.

This last one is the one that happened to me. I wasn't the ex-boyfriend (thank God). I was the friend who was told about it. I took it very very seriously, investigated as best I could, and eventually became extremely confident that the accusation was false. No, you don't know the people involved. No, I won't give you personal details. No, I won't tell you how I became certain that the accusation was false because that would involve personal details. Yes, that leaves you a lot of room to accuse me of lying if you want.

But if my word isn't good enough for you, I happen to have witnessed two *more* cases of false rape accusations where I *can* tell you some minimal details. In a psychiatric hospital I

used to work in (not the one I currently work in) during my brief time there there were two different accusations of rape by staff members against patients…

I want to take a second out to say *very emphatically* that *all accusations of rape by psychiatric patients should be taken very seriously*. Yes, psychiatric patients sometimes have complicated cognitive or personality issues that make them more likely to falsely report rape, but for *exactly this reason* they are much more vulnerable and people are much more likely to take advantage of them. This is a *known problem* and you should *never dismiss their complaint*.

…but in this case, there were video cameras all over the hospital and these were sufficient to prove that no assault had taken place in either case. Now I know someone is going to say that blah blah psychiatric patients blah blah doesn't generalize to the general population, but the fact is that even if you accept that sorta-ableist dismissal, those patients were in hospital for three to seven days and then they went back out into regular society.I would love to say that we treated every single one of their problems so thoroughly it would never come back but I wouldn't bet on it.

So I know three men who have been accused of rape in a way that did not involve the police, and none (as far as I know) who have been accused in a way that did. This suggests that like rapes themselves, most false rape accusations never reach law enforcement.

While rape victims have some incentives to report their cases to the police – a desire for justice, a desire for safety, the belief that the evidence will support them – false accusers have very strong incentives not to – too much work, easier revenge through other means, knowledge that the evidence is unlikely

to support them, fear of getting in trouble for perjury if their deception gets out. So I consider it a very conservative estimate to say that the ratio of unreported to reported false accusations is 4:1 – the same as it is with rapes. A more realistic estimate might be as high as double or triple that.

V.

Now we have the data necessary to do a slightly better job calculating the risk of false rape allegations. We'll start with the most conservative possible estimate.

We will stick with the article's figure of 84,000 reported rapes per year and 8% false accusation rate, for a total of 6,750 falsely accused.

We go on to assume, for the sake of conservativism, that there has never been a single false accuser who did not later confess, and that there has never been a false accuser who did not go to the police (my own memories of this must be hallucinations).

Since there are 53 million men ages 15-39 in the United States, the probability of being one of these 6,750 falsely accused is 1/7850 per year. But since you have 24 years in that age range in which to be accused, your lifetime probability of being falsely accused is about 1/327, or 0.3%. This is small, but according to Clymer's list it's about the same as your risk of dying in a car crash. Do you worry about dying in a car crash? Then you are allowed to worry about being falsely accused of rape.

(note that this is the most conservative possible estimate, using exactly the same numbers as in the article but not lying about what math we're doing. But the article got 1/750,000. So the absolute lower bound for how wrong the article was is "wrong by a factor of 2,300x")

What about a slightly less hyperconservative estimate? Continuing our conservative assumption that there has never been a false accuser who has not later confused, but allowing that false accusations reach the police at only the same rate that rapes do, 1.5% of men will get falsely accused.

What estimate do I personally find most likely? Suppose we keep everything else the same, but allow that for every false accuser who later confesses, there is also one false accuser who does not later confess. This raises the false accusation rate to 16% – which, keep in mind, is still less than half of what the police think it is, so it's not like we're allowing rape-culture-happy cops to color our perception here. Now 3% of men will get falsely accused.

What is an upper bound for the extent of this problem? We could obtain one by using Kanin's 40% and holding everything else constant, but no matter how many times I qualified this attempt with "I am using this as an upper bound, not endorsing this as the actual number of rapes", someone would yell at me for using a study they disagree with and call me a rape apologist. So I will leave the difficult task of multiplying 3% by 2.5x to my readers. You might then try multiplying it even further if you think false accusations are less likely than true accusations to make it to the police.

So greater than 0.3% of men get falsely accused of rape sometime in their lives, and the most likely number is probably around 3%.

Which means the article was off by a factor of at least 2,300x and probably more like 22,700x.

And yet it got 35,000 Tumblr likes and reblogs. By blatantly lying in a sensationalist way, it became more popular than anything you or I will ever write. There are scientists

dedicating their lives to making new discoveries on the frontiers of knowledge, poets making words dance and catch fire, struggling writers trying to tell the stories inside of them – all desperate for someone to pay attention to what they're saying – and the Internet ignores these people and instead brings hundreds of thousands of hits and no doubt a big windfall in ad revenue to frickin' Buzzfeed.

And I would like to just let it be, except that there's a probably one-in-thirty but definitely-no-less-than-one-in-three-hundred chance that I will be falsely accused of rape someday, and need to defend myself, and maybe I'll have what should be an airtight alibi, and then the people who read this Buzzfeed article will dismiss it with "Well, I saw on the Internet there's only a one in a million chance you're telling the truth, so screw your alibi!" This is *already happening*. One of the Tumblr rebloggers added the comment "Yeah, so you know the dude who says he was falsely accused of rape? Now you know. He's a rapist." These are not just falsehoods, they're *dangerous* falsehoods.

So please permit me a second to gripe about this.

It is commonly said that a lie will get halfway across the world before the truth can get its boots on. And this is true. Except in the feminist blogosphere, where a lie will get to Alpha Centauri and back three times while the truth is locked up in a makeshift dungeon in the basement, screaming.

I have been debunking bad statistics for a long time. In medicine, in psychology, in politics. Click on the "statistics" tag of this blog if you don't believe me. Yet the feminist blogosphere is the *only* place where I consistently see things atrociously wrong get reblogged by thousands of usually very smart people without anyone ever bothering to think critically

about them. Like, thirty five thousand feminists – including some who self-identify as rationalists! – saw an article that literally said a guy was *more likely to get hit by a comet than get falsely accused of rape*, and said "Yeah, sure, that sounds plausible".

So please permit me to keep griping just one moment longer. *Be extraordinarily paranoid when dealing with the feminist blogosphere.* This may be true of all highly charged political blogospheres, but it is *certainly* true of feminism. If you go in there with an innocent attitude of "Here is a number, I assume it is generally correct and means what it says it means", you will get *super-burned*

There are some honorable exceptions. I have found Alas, A Blog to be pretty scrupulous, and of course everything ever written by Ozy is wonderful and perfect in every way. But two swallows do not make a summer, and these and any similar blogs you find should be considered islands of lucidity battered by a constant tide of bullshytte. I do not have time to debunk them all but you should view them with a prior of extraordinarily high suspicion.

Thank you for letting me get that out of my system.

VI.

Why would this happen? Why would smart people, by the tens of thousands, be so delighted by the opportunity to embrace these fabrications?

There is something called the "just world fallacy", that says everyone gets what they deserve and moral questions are always easy and there is never any need to make scary tradeoffs.

And, as is so often the case for things with “fallacy” in the name, it is not true.

Look at how the Clymer article, in its own words, describes false rape allegations:

“False rape hysteria”, it informs us, is perpetrated by “men’s rights activists, more accurately known as insecure woman-hating assholes”, because they think “women are products to be bought and sold and when these objects assert their right to human value many (if not most) men feel threatened.”

Now let’s hear from a guy on the r/mensrights community on Reddit:

> Anyway, like I said, it’s been just over a year since [I was falsely accused of rape]. Since then I haven’t been the same. The most striking thing that I’ve noticed is the paranoia that I have almost every waking moment. Of everybody. Of men, of women, and even friends. I can’t bring myself to date women anymore. I have panic attacks every time I see a police officer. I constintly think that I’m being followed. The night I came home from being interviewed by the cops I drank myself to sleep and I’ve been doing that ever since. If I don’t any flicker of light makes me think that the police are here to arrest me. I’ve been able to fake a normal social life to my family and work and the friends I have left but most don’t know anything about this. I’m not looking for pity from anyone. In fact, I’m doing better than I have been. The reason I’m posting this is because I want people to know how bad being accused of something like rape can hurt and scar someone.

Man, what an “insecure, woman-hating asshole.”

But consider the alternative to this kind of glib dismissal.

3% of men are falsely accused of rape. 15% of women are raped. If someone you know gets accused of rape, your prior still is very very high that they did it.

I was extraordinarily lucky to find very strong evidence that my friend was innocent. I was extraordinarily lucky that both my co-workers had video feeds that could confirm their stories. If I hadn't, I don't know what I would have done. My two choices would have been to either accept the possibility that I'm staying friends with a rapist, or to accept the possibility I'm ostracizing someone for something he didn't do.

And someone is going to expect me to conclude by recommending what the correct thing to do in these cases is, but *I have no idea*. Probably there is no solution that isn't horrible. If there is, it's way above my pay grade. Ask Ozy. Ze's the one with the Gender Studies degree.

All I can suggest is that you not flee from the magnitude of the decision with comfortable lies.

One of those comfortable lies is to tell yourself that all women are lying sluts so the accusation can be safely ignored.

But another comfortable lie is that false rape accusations are eleven times rarer than getting hit by comets.

This is why a terrible article on Buzzfeed is getting more publicity and support than anything you or I will ever write.

Because people want to live in his world, where the comfortable lies are all true and no one suffers without deserving it.

In Favor of Niceness, Community, and Civilization

[Content warning: Discussion of social justice, discussion of violence, spoilers for Jacqueline Carey books.]

[Edit 10/25: This post was inspired by a debate with a friend of a friend on Facebook who has since become somewhat famous. Although I strongly disagree with him on the point at issue here, I have nothing against him personally. Since some people have (ironically) been using this post to attack him every time he says anything at all, I have decided to obfuscate his identity under the pseudonym "Andrew Cord" in order to make this a little harder.]

I.

Andrew Cord criticizes me for my bold and controversial suggestion that maybe people should try to tell slightly fewer blatant hurtful lies:

> I just find it kind of darkly amusing and sad that the "rationalist community" loves "rationality is winning" so much as a tagline and yet are clearly not winning. And then complain about losing rather than changing their tactics to match those of people who are winning.
>
> Which is probably because if you *really* want to be the kind of person who wins you have to actually care about winning something, which means you have to have politics, which means you have to embrace "politics the mindkiller" and "politics is war and arguments are soldiers", and Scott would clearly rather spend the rest of his life losing than do this.

> That post [the one debunking false rape statistics] is exactly my problem with Scott. He seems to honestly think that it's a worthwhile use of his time, energy and mental effort to download evil people's evil worldviews into his mind and try to analytically debate them with statistics and cost-benefit analyses.
>
> He gets *mad* at people whom he detachedly intellectually agrees with but who are willing to back up their beliefs with war and fire rather than pussyfooting around with debate-team nonsense.
>
> It honestly makes me kind of sick. It is exactly the kind of thing that "social justice" activists like me *intend* to attack and "trigger" when we use "triggery" catchphrases about the mewling pusillanimity of privileged white allies.

In other words, if a fight is important to you, fight nasty. If that means lying, lie. If that means insults, insult. If that means silencing people, silence.

It always makes me happy when my ideological opponents come out and say eloquently and openly what I've always secretly suspected them of believing. It's even better when the person involved is a celebrity, and I can tell people "Hey! I argued with a celebrity!"

My natural instinct is to give some of the reasons why I think Andrew is wrong, starting with the history of the "noble lie" concept and moving on to some examples of why it didn't work very well, and why it might not be expected not to work so well in the future.

But in a way, that would be assuming the conclusion. I wouldn't be showing respect for Andrew's arguments. I

wouldn't be going halfway to meet them on their own terms.

The respectful way to rebut Andrew's argument would be to spread malicious lies about Andrew to a couple of media outlets, fan the flames, and wait for them to destroy his reputation.

Then if the stress ends up bursting an aneurysm in his brain, I can dance on his grave, singing:

> ♪ ♬ I won this debate in a very effective manner. Now you can't argue in favor of nasty debate tactics any more ♬ ♪

I am not going to do that, but if I *did* it's unclear to me how Andrew could object. I mean, he thinks that sexism is detrimental to society, so spreading lies and destroying people is justified in order to stop it. I think that discourse based on mud-slinging and falsehoods is detrimental to society. Therefore…

II.

But really, all this talk of lying and spreading rumors about people is – what was Andrew's terminology – "pussyfooting around with debate-team nonsense". You know who got things done? The IRA. They didn't agree with the British occupation of Northern Ireland and they weren't afraid to let people know in that very special way only a nail-bomb shoved through your window at night can.

Why not assassinate prominent racist and sexist politicians and intellectuals? I won't name names since that would be crossing a line, but I'm sure you can generate several of them who are sufficiently successful and charismatic that, if knocked off, there would not be an equally competent racist or sexist

immediately available to replace them, and it would thus be a serious setback for the racism/sexism movement.

Other people can appeal to "the social contract" or "the general civilizational rule not to use violence", but not Andrew:

> I think that whether or not I use certain weapons has zero impact on whether or not those weapons are used against me, and people who think they do are either appealing to a kind of vague Kantian morality that I think is invalid or a specific kind of "honor among foes" that I think does not exist.

And don't give me that nonsense about the police. I'm sure a smart person like you can think of clever exciting new ways to commit the perfect murder. Unless you do not believe there will *ever* be an opportunity to defect unpunished, you need this sort of social contract to take you at least some of the way.

He continues:

> When Scott calls rhetorical tactics he dislikes "bullets" and denigrates them it actually hilariously plays right into this point…to be "pro-bullet" or "anti-bullet" is ridiculous. Bullets, as you say, are neutral. I am in favor of my side using bullets as best they can to destroy the enemy's ability to use bullets.
>
> In a war, a real war, a war for survival, you use all the weapons in your arsenal because you assume the enemy will use all the weapons in theirs. Because you understand that it IS a war.

There are a lot of things I am tempted to say to this.

Like “And that is why the United States immediately nukes every country it goes to war with.”

Or “And that is why the Geneva Convention was so obviously impossible that no one even bothered to attend the conference”.

Or “And that is why, to this very day, we solve every international disagreement through total war.”

Or “And that is why Martin Luther King was immediately reduced to a nonentity, and we remember the Weathermen as the sole people responsible for the success of the civil rights movement”

But I think what I am *actually* going to say is that, for the love of God, if you like bullets so much, stop using them as a metaphor for ‘spreading false statistics’ and go buy a gun.

(I just realized I probably shouldn’t say that. If I get shot in the next while, someone point the police here.)

III.

So let’s derive why violence is not in fact The One True Best Way To Solve All Our Problems. You can get most of this from Hobbes, but this blog post will be shorter.

Suppose I am a radical Catholic who believes all Protestants deserve to die, and therefore go around killing Protestants. So far, so good.

Unfortunately, there might be some radical Protestants around who believe all Catholics deserve to die. If there weren’t before, there probably are now. So they go around killing Catholics, we’re both unhappy and/or dead, our economy tanks, hundreds of innocent people end up as collateral damage, and our country goes down the toilet.

So we make an agreement: I won't kill any more Catholics, you don't kill any more Protestants. The specific Irish example was called the Good Friday Agreement and the general case is called "civilization".

So then I try to destroy the hated Protestants using the government. I go around trying to pass laws banning Protestant worship and preventing people from condemning Catholicism.

Unfortunately, maybe the next government in power is a Protestant government, and they pass laws banning Catholic worship and preventing people from condemning Protestantism. No one can securely practice their own religion, no one can learn about other religions, people are constantly plotting civil war, academic freedom is severely curtailed, and once again the country goes down the toilet.

So again we make an agreement. I won't use the apparatus of government against Protestantism, you don't use the apparatus of government against Catholicism. The specific American example is the First Amendment and the general case is called "liberalism", or to be dramatic about it, "civilization 2.0"

Every case in which both sides agree to lay down their weapons and be nice to each other has corresponded to spectacular gains by both sides and a new era of human flourishing.

"Wait a second, no!" someone yells. "I see where you're going with this. You're going to say that agreeing not to spread malicious lies about each other would also be a civilized and beneficial system. Like maybe the Protestants could stop saying that the Catholics worshipped the Devil, and the Catholics could stop saying the Protestants hate the Virgin

Mary, and they could both relax the whole thing about the Jews baking the blood of Christian children into their matzah.

“But your two examples were about contracts written on paper and enforced by the government. So maybe a ‘no malicious lies’ amendment to the Constitution would work if it were enforceable, *which it isn’t*, but just *asking* people to stop spreading malicious lies is doomed from the start. The Jews will no doubt spread lies against *us*, so if we stop spreading lies about them, all we’re doing is abandoning an effective weapon against a religion I personally know to be heathenish! Rationalists should win, so put the blood libel on the front page of every newspaper!”

Or, as Andrew puts it:

> Whether or not I use certain weapons has zero impact on whether or not those weapons are used against me, and people who think they do are either appealing to a kind of vague Kantian morality that I think is invalid or a specific kind of “honor among foes” that I think does not exist.

So let’s talk about how beneficial game-theoretic equilibria can come to exist even in the absence of centralized enforcers. I know of two main ways: reciprocal communitarianism, and divine grace.

Reciprocal communitarianism is probably how altruism evolved. Some mammal started running TIT-FOR-TAT, the program where you cooperate with anyone whom you expect to cooperate with you. Gradually you form a successful community of cooperators. The defectors either join your community and agree to play by your rules or get outcompeted.

Divine grace is more complicated. I was tempted to call it "spontaneous order" until I remembered the rationalist proverb that if you don't understand something, you need to call it by a term that reminds you that don't understand it or else you'll think you've explained it when you've just named it.

But consider the following: I am a pro-choice atheist. When I lived in Ireland, one of my friends was a pro-life Christian. I thought she was responsible for the unnecessary suffering of millions of women. She thought I was responsible for killing millions of babies. And yet she invited me over to her house for dinner without poisoning the food. And I ate it, and thanked her, and sent her a nice card, without smashing all her china.

Please try not to be insufficiently surprised by this. Every time a Republican and a Democrat break bread together with good will, it is a miracle. It is an equilibrium as beneficial as civilization or liberalism, which developed in the total absence of any central enforcing authority.

When you look for these equilibria, there are lots and lots. Andrew says there is no "honor among foes", but if you read the *Iliad* or any other account of ancient warfare, there is practically nothing *but* honor among foes, and it wasn't generated by some sort of Homeric version of the Geneva Convention, it just sort of happened. During World War I, the English and Germans spontaneously got out of their trenches and celebrated Christmas together with each other, and on the sidelines Andrew was shouting "No! Stop celebrating Christmas! Quick, shoot them before they shoot you!" but they didn't listen.

All I will say in way of explaining these miraculous equilibria is that they seem to have something to do with inheriting a cultural norm and not screwing it up. Punishing the occasional

defector seems to be a big part of not screwing it up. How exactly that cultural norm came to be is less clear to me, but it might have something to do with the reasons why an entire civilization's bureaucrats may suddenly turn 100% honest at the same time. I'm pretty sure I'm supposed to say the words timeless decision theory around this point too, and perhaps bring up the kind of Platonic contract that I have written about previously.

I think most of our useful social norms exist through a combination of divine grace and reciprocal communitarianism. To some degree they arise spontaneously and are preserved by the honor system. To another degree, they are stronger or weaker in different groups, and the groups that enforce them are so much more pleasant than the groups that don't that people are willing to go along.

The norm against malicious lies follows this pattern. Politicians lie, but not *too much*. Take the top story on Politifact Fact Check today. Some Republican claimed his supposedly-maverick Democratic opponent actually voted with Obama's economic policies 97 percent of the time. Fact Check explains that the statistic used was actually for *all* votes, not just economic votes, and that members of Congress typically have to have >90% agreement with their president because of the way partisan politics work. So it's a lie, and is properly listed as one. But it's a lie based on slightly misinterpreting a real statistic. He didn't just totally make up a number. He didn't even just make up something else, like "My opponent personally helped design most of Obama's legislation".

Even Clymer lied less than he *possibly could have*. He got his fake numbers by conflating rapes per sex act with rapes per lifetime, and it's really hard for me to imagine someone doing

that by anything resembling accident. But he couldn't bring himself to go the extra step and just totally make up numbers with no grounding whatsoever. And part of me wonders: why not? If you're going to use numbers you know are false to destroy people, why is it better to derive the numbers through a formula you know is incorrect, than to just skip the math and make the numbers up in the first place? "The FBI has determined that no false rape claims have ever been submitted, my source is an obscure report they published, when your local library doesn't have it you will just accept that libraries can't have all books, and suspect nothing."

This would have been a *more believable* claim than the one he made. Because he showed his work, it was easy for me to debunk it. If he had just said it was in some obscure report, I wouldn't have gone through the trouble. So why did he go the harder route?

People *know* lying is wrong. They know if they lied they would be punished. More ~~spontaneous social order~~ miraculous divine grace. And so they want to hedge their bets, be able to say "Well, I didn't exactly *lie*, per se."

And this is good! We *want* to make it politically unacceptable to have people say that Jews bake the blood of Christian children into their matzah. Now we build on that success. We start hounding around the edges of currently acceptable lies. "Okay, you didn't *literally* make up your statistics, but you still lied, and you still should be cast out from the community of people who have reasonable discussions and never trusted by anyone again."

It might not totally succeed in making a new norm against this kind of thing. But at least it will prevent other people from

seeing Clymer's success, taking heart, and having the number of lies which are socially acceptable gradually *advance*.

So much for protecting what we have been given by divine grace. But there is also reciprocal communitarianism to think of.

I seek out people who signal that they want to discuss things honestly and rationally. Then I try to discuss things honestly and rationally with those people. I try to concentrate as much of my social interaction there as possible.

So far this project is going pretty well. My friends are nice, my romantic relationships are low-drama, my debates are productive and I am learning so, so much.

And people think "Hm, I could hang out at 4Chan and be called a 'fag'. Or I could hang out at Slate Star Codex and discuss things rationally and learn a lot. And if I want to be allowed in, all I have to do is not be an intellectually dishonest jerk."

And so our community grows. And all over the world, the mysterious divine forces favoring honest and kind equilibria gain a little bit more power over the mysterious divine forces favoring lying and malicious equilibria.

Andrew thinks I am trying to fight all the evils of the world, and doing so in a stupid way. But sometimes I just want to cultivate my garden.

IV.

Andrew goes on to complain:

> Scott…seems to [dispassionately debate] evil people's evil worldviews …with statistics and cost-benefit analyses.

> He gets *mad* at people whom he detachedly intellectually agrees with but who are willing to back up their beliefs with war and fire rather than pussyfooting around with debate-team nonsense.

I accept this criticism as an accurate description of what I do.

Compare to the following two critiques: "The Catholic Church wastes so much energy getting upset about heretics who believe *mostly* the same things as they do, when there are literally *millions* of Hindus over in India who don't believe in Catholicism *at all*! What dumb priorities!"

Or "How could Joseph McCarthy get angry about a couple of people who *might* have been Communists in the US movie industry, when over in Moscow there were *thousands* of people who were openly *super* Communist *all the time*?"

There might be foot-long giant centipedes in the Amazon, but I am a lot more worried about boll weevils in my walled garden.

Creationists lie. Homeopaths lie. Anti-vaxxers lie. This is part of the Great Circle of Life. It is not necessary to call out every lie by a creationist, because the sort of person who is still listening to creationists is not the sort of person who is likely to be moved by call-outs. There is a role for *organized* action against creationists, like preventing them from getting their opinions taught in schools, but the marginal blog post "debunking" a creationist something something is a waste of time. Everybody who wants to discuss things rationally has already formed a walled garden and locked the creationists outside of it.

Anti-Semites fight nasty. The Ku Klux Klan fights nasty. Neo-Nazis fight nasty. We dismiss them with equanamity, in

accordance with the ancient proverb: “Haters gonna hate”. There is a role for *organized* opposition to these groups, like making sure they can’t actually terrorize anyone, but the marginal blog post condemning Nazism is a waste of time. Everybody who wants to discuss things charitably and compassionately has already formed a walled garden and locked the Nazis outside of it.

People who want to discuss things rationally and charitably have not yet locked Charles Clymer out of their walled garden.

He is not a heathen, he is a heretic. He is not a foreigner, he is a traitor. He comes in talking all liberalism and statistics, and then he betrays the signals he has just sent. He is not just some guy who defects in the Prisoner’s Dilemma. He is the guy who defects while wearing the “I COOPERATE IN PRISONERS DILEMMAS” t-shirt.

What really, *really* bothered me wasn’t Clymer at all: it was that *rationalists* were taking him seriously. Smart people, kind people! I even said so in my article. Boll weevils in our beautiful walled garden!

Why am I always harping on feminism? I feel like we’ve got a good thing going, we’ve ratified our Platonic contract to be intellectually honest and charitable to each other, we are going about perma-cooperating in the Prisoner’s Dilemma and reaping gains from trade.

And then someone says “Except that of course regardless of all that I reserve the right to still use lies and insults and harassment and dark epistemology to spread feminism”. Sometimes they do this explicitly, like Andrew did. Other times they use a more nuanced argument like “Surely you didn’t think the same rules against lies and insults and harassment should apply to oppressed and privileged people,

did you?" And other times they don't say anything, but just show their true colors by reblogging an awful article with false statistics.

(and still other times they don't do any of this and they are wonderful people whom I am glad to know)

But then someone else says "Well, if they get their exception, I deserve my exception," and then someone else says "Well, if those two get exceptions, I'm out", and *you have no idea how difficult it is to successfully renegotiate the terms of a timeless Platonic contract that doesn't literally exist*.

No! I am Exception Nazi! NO EXCEPTION FOR YOU! Civilization didn't conquer the world by forbidding you to murder your enemies *unless* they are actually unrighteous in which case go ahead and kill them all. Liberals didn't give their lives in the battle against tyranny to end discrimination against all religions *except* Jansenism because seriously fuck Jansenists. Here we have built our Schelling fence and here we are defending it to the bitter end.

V.

Contrary to how it may appear, I am not trying to doom feminism.

Feminists like to mock the naivete of anyone who says that classical liberalism would suffice to satisfy feminist demands. And true, you cannot simply assume Adam Smith and derive Andrea Dworkin. Not being an asshole to women and not writing laws declaring them officially inferior are both good starts, but it not enough if there's still cultural baggage and entrenched gender norms.

But here I am, defending this principle – kind of a supercharged version of liberalism – of "It is not okay to use

lies, insults, and harassment against people, even if it would help you enforce your preferred social norms."

And I notice that this gets us a heck of a lot closer to feminism than Andrew's principle of "Go ahead and use lies, insults, and harassment if they are effective ways to enforce your preferred social norms."

Feminists are very concerned about slut-shaming, where people harass women who have too much premarital sex. They point out that this is very hurtful to women, that men might underestimate the amount of hurt it causes women, and that the standard-classical-liberal solution of removing relevant government oppression does nothing. All excellent points.

But one assumes the harassers think that women having premarital sex is detrimental to society. So they apply their general principle: "I should use lies, insults, and harassment to enforce my preferred social norms."

But this is the principle Andrew is asserting, against myself and liberalism.

Feminists think that women should be free from fear of rape, and that, if raped, no one should be able to excuse themselves with "well, she was asking for it".

But this is the same anti-violence principle as saying that the IRA shouldn't throw nail-bombs through people's windows or that, nail bombs having been thrown, the IRA can't use as an excuse "Yeah, well, they were complicit with the evil British occupation, they deserved it." Again, I feel like I'm defending this principle a whole lot more strongly and consistently than Andrew is.

Feminists are, shall we say, divided about transgender people, but let's allow that the correct solution is to respect their

rights.

When I was young and stupid, I used to believe that transgender was really, really dumb. That they were looking for attention or making it up or something along those lines.

Luckily, since I was a classical liberal, my reaction to this mistake was – to not bother them, and to get very very angry at people who did bother them. I got upset with people trying to fire Phil Robertson for being homophobic even though homophobia is stupid. You better bet I also got upset with people trying to fire transgender people back when I thought transgender was stupid.

And then I grew older and wiser and learned – hey, transgender isn't stupid at all, they have very important reasons for what they do and go through and I was atrociously wrong. And I said a mea culpa.

But it could have been worse. I didn't like transgender people, and so I *left them alone while still standing up for their rights*. My epistemic structure *failed gracefully*. For anyone who's not overconfident, and so who expects massive epistemic failure on a variety of important issues all the time, graceful failure modes are a *really important feature* for an epistemic structure to have.

God only knows what Andrew would have done, if through bad luck he had accidentally gotten it into his head that transgender people are bad. From his own words, we know he wouldn't be "pussyfooting around with debate-team nonsense".

I admit there are many feminist principles that cannot be derived from, or are even opposed to my own liberal principles. For example, some feminists have suggested that pornography be banned because it increases the likelihood of

violence against women. Others suggest that research into gender differences should be banned, or at least we should stigmatize and harass the researchers, because any discoveries made might lend aid and comfort to sexists.

To the first, I would point out that there is now strong evidence that pornography, especially violent objectifying pornography, very significantly decreases violence against women. I would ask them whether they're happy that we did the nice liberal thing and waited until all the evidence came in so we could discuss it rationally, rather than immediately moving to harass and silence anyone taking the pro-pornography side.

And to the second, well, we have a genuine disagreement. But I wonder whether they would prefer to discuss that disagreement reasonably, or whether we should both try to harass and destroy the other until one or both of us are too damaged to continue the struggle.

And if feminists agree to have that reasonable discussion, but lose, I would tell them that they get a consolation prize. Having joined liberal society, they can be sure that no matter what those researchers find, I and all of their new liberal-society buddies will fight tooth and nail against anyone who uses any tiny differences those researchers find to challenge the central liberal belief that everyone of every gender has basic human dignity. Any victory for me is going to be a victory for feminists as well; maybe not a perfect victory, but a heck of a lot better than what they have right now.

VI.

I am not trying to fight all the evils of the world. I am just trying to cultivate my garden.

And you argue: "But isn't that selfish and oppressive and privileged? Isn't that confining everyone outside of your

walled garden to racism and sexism and nastiness?

But there is a famous comic which demonstrates what can happen to certain walled gardens.

Why yes, it does sound like I'm making the unshakeable assumption that liberalism always wins, doesn't it? That people who voluntarily relinquish certain forms of barbarism will be able to gradually expand their territory against the hordes outside, instead of immediately being conquered by their less scrupulous neighbors? And it looks like Andrew isn't going to let that assumption pass.

He writes:

> The *whole history* of why the institutional Left in our society is a party of toothless, spineless, gutless losers and they've spent two generations doing nothing but lose.

One is reminded of the old joke about the Nazi papers. The rabbi catches an old Jewish man reading the Nazi newspaper and demands to know how he could look at such garbage. The man answers "When I read our Jewish newpapers, the news is so depressing – oppression, death, genocide! But here, everything is great! We control the banks, we control the media. Why, just yesterday they said we had a plan to kick the Gentiles out of Germany entirely!"

And I have two thoughts about this.

First, it argues that "Evil people are doing evil things, so we are justified in using any weapons we want to stop them, no matter how nasty" suffers from a certain flaw. Everyone believes their enemies are evil people doing evil things. If you're a Nazi, you are just defending yourself, in a very proportionate manner, against the Vast Jewish Conspiracy To Destroy All Germans.

But second, before taking Andrew's words for how disastrously liberalism is doing, we should check the newspapers put out by liberalism's enemies. Here's Mencius Moldbug:

> Cthulhu may swim slowly. But he only swims left. Isn't that interesting?
>
> In each of the following conflicts in Anglo-American history, you see a victory of left over right: the English Civil War, the so-called "Glorious Revolution," the American Revolution, the American Civil War, World War I, and World War II. Clearly, if you want to be on the winning team, you want to start on the left side of the field.
>
> Where is the John Birch Society, now? What about the NAACP? Cthulhu swims left, and left, and left. There are a few brief periods of true reaction in American history – the post-Reconstruction era or Redemption, the Return to Normalcy of Harding, and a couple of others. But they are unusual and feeble compared to the great leftward shift. McCarthyism is especially noticeable as such. And you'll note that McCarthy didn't exactly win.
>
> In the history of American democracy, if you take the mainstream political position (Overton Window, if you care) at time T1, and place it on the map at a later time T2, T1 is always way to the right, near the fringe or outside it. So, for instance, if you take the average segregationist voter of 1963 and let him vote in the 2008 election, he will be way out on the wacky right wing. Cthulhu has passed him by.

I've got to say Mencius makes a much more convincing argument than Andrew does.

Robert Frost says "A liberal is a man too broad-minded to take his own side in a quarrel". Ha ha ha.

And yet, outside of Saudi Arabia you'll have a hard time finding a country that doesn't at least pay lip service to liberal ideas. Stranger still, many of those then go on to *actually implement them*, either voluntarily or after succumbing to strange pressures they don't understand. In particular, the history of the past few hundred years in the United States has been a history of decreasing censorship and increasing tolerance.

Contra the Reactionaries, feminism isn't an exception to that, it's a casualty of it. 1970s feminists were saying that all women need to rise up and smash the patriarchy, possibly with literal smashing-implements. 2010s feminists are saying that if some women want to be housewives, that's great and their own choice because in a liberal society everyone should be free to pursue their own self-actualization.

And that has *corresponded to* spectacular successes of the specific causes liberals like to push, like feminism, civil rights, gay marriage, et cetera, et cetera, et cetera.

A liberal is a man too broad-minded to take his own side in a quarrel. And yet when liberals enter quarrels, they always win. Isn't that interesting?

VII.

Andrew thinks that liberals who voluntarily relinquish any form of fighting back are just ignoring perfectly effective weapons. I'll provide the quote:

> In a war, a real war, a war for survival, you use all the weapons in your arsenal because you assume the enemy will use all the weapons in theirs. Because you understand that it IS a war… Any energy spent mentally debating how, in a perfect world run by a Lawful Neutral Cosmic Arbiter that will never exist, we could settle wars without bullets is energy you could better spend down at the range improving your marksmanship… I am amazed that the "rationalist community" finds it to still be so opaque.

Let me name some other people who mysteriously managed to miss this perfectly obvious point.

The early Christian Church had the slogan "resist not evil" (Matthew 5:39), and indeed, their idea of Burning The Fucking System To The Ground was to go unprotestingly to martyrdom while publicly forgiving their executioners. They were up against the Roman Empire, possibly the most effective military machine in history, ruled by some of the cruelest men who have ever lived. By Andrew's reckoning, this should have been the biggest smackdown in the entire history of smackdowns.

And it kind of was. Just not the way most people expected.

Mahatma Gandhi said "Non-violence is the greatest force at the disposal of mankind. It is mightier than the mightiest weapon of destruction devised by the ingenuity of man." Another guy who fought one of the largest empires ever to exist and won resoundingly. And he was pretty insistent on truth too: "Non-violence and truth are inseparable and presuppose one another."

Also skilled at missing the obvious: Martin Luther King. Desmond Tutu. Aung San Suu Kyi. Nelson Mandela was

smart and effective at the beginning of his career, but fell into a pattern of missing the obvious when he was older. Maybe it was Alzheimers.

Of course, there are counterexamples. Jews who nonviolently resisted the Nazis didn't have a very good track record. You need a certain pre-existing level of civilization for liberalism to be a good idea, and a certain pre-existing level of liberalism for supercharged liberalism where you don't spread malicious lies and harass other people to be a good idea. You need to have pre-existing community norms in place before trying to summon mysterious beneficial equilibria.

So perhaps I am being too harsh on Andrew, to contrast him with Aung San Suu Kyi and her ilk. After all, all Aung San Suu Kyi had to do was fight the Burmese junta, a cabal of incredibly brutal military dictators who killed several thousand people, tortured anyone who protested against them, and sent eight hundred thousand people they just didn't like to forced labor camps. Andrew has to deal with *people who aren't as feminist as he is*. Clearly this requires much stronger measures!

VIII.

Liberalism does not conquer by fire and sword. Liberalism conquers by communities of people who agree to play by the rules, slowly growing until eventually an equilibrium is disturbed. Its battle cry is not "Death to the unbelievers!" but "If you're nice, you can join our cuddle pile!"

(I have been to New York Less Wrong meetups, and know that this is also effective when meant literally)

But some people, through lack of imagination, fail to find this battle cry sufficiently fear-inspiring.

I hate to invoke fictional evidence, especially since perhaps Andrew's strongest point is that the real world doesn't work like fiction. But these people need to read Jacqueline Carey's *Kushiel's Avatar*.

Elua is the god of kindness and flowers and free love. All the other gods are gods of blood and fire, and Elua is just like "Love as thou wilt" and "All knowlege is worth having". He is the patron deity of exactly the kind of sickeningly sweet namby-pamby charitable liberalism that Andrew is complaining about.

And there is a certain commonality to a lot of the Kushiel books, where some tyrant or sorcerer thinks that a god of flowers and free love will be a pushover, and starts harassing his followers. And the only Eluite who shows up to stop him is Phèdre nó Delaunay, and the tyrant thinks "Ha! A woman, who doesn't even know how to fight, doesn't have any magic! What a wuss!"

But here is an important rule about dealing with fantasy book characters.

If you ever piss off Sauron, you should probably find the Ring of Power and take it to Mount Doom.

If you ever get piss off Voldemort, you should probably start looking for Horcruxes.

If you ever piss off Phèdre nó Delaunay, *run and never stop running*.

Elua is the god of flowers and free love and he is terrifying. If you oppose him, there will not be enough left of you to bury, and it will not matter because there will not be enough left of your city to bury you in.

And Jacqueline Carey and Mencius Moldbug are both wiser than Andrew Cord.

Carey portrays liberalism as Elua, a terrifying unspeakable Elder God who is fundamentally good.

Moldbug portrays liberalism as Cthulhu, a terrifying unspeakable Elder God who is fundamentally evil.

But Andrew? He *doesn't even seem to realize liberalism is a terrifying unspeakable Elder God at all*. It's like, *what?*

Andrew is the poor shmuck who is sitting there saying "Ha ha, a god who doesn't even control any hell-monsters or command his worshippers to become killing machines. What a weakling! This is going to be so easy!"

And you want to scream: "THERE IS ONLY ONE WAY THIS CAN POSSIBLY END AND IT INVOLVES YOU BEING EATEN BY YOUR OWN LEGIONS OF DEMONAICALLY CONTROLLED ANTS"

(uh, spoilers)

XII. Politicization

Right is the New Left

[Content warning: some ideas that might make you feel anxious about your political beliefs. Epistemic status: very speculative and not necessarily endorsed. This post is less something I will defend to the death and more a form of self-therapy.]

I.

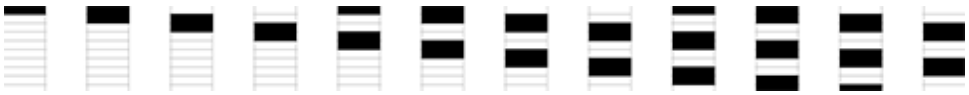
Let's explain fashion using cellular automata. This isn't going to be cringe-inducingly nerdy at all!

We'll start with a one-dimensional vertical "world" a single cell thick and however many cells we want tall. Cells can be in one of two states, "black" or "white". We start with the top cell "black" and all the other cells white, and the world changes with granular time ("ticks") according to the following rules:

1. On each tick, a cell tries to be the same color that the cell above it was last tick.
2. On each tick, a cell tries NOT to be the same color that the cell below it was last tick.
3. If they ever conflict, Rule 1 takes precedence over Rule 2.
4. If none of these rules apply, a cell stays as it is.

Here's what we get with a world four cells tall.

And here's what we get with a world ten cells tall.



It looks like what we're getting is a "setup period" as the column fills, followed by a "sandwich effect" of two-cell-tall black rectangles separated by two-cell-tall white rectangles gradually moving down the column. Although this isn't *really* what happens, it also looks like rectangles that fall off the bottom reappear on the top. The overall effect is sort of like a barber pole.

Okay, now let's get to the fashion.

Consider a group of people separated by some ranked attribute. Let's call it "class". There are four classes: the upper class, the middle class, the lower class, and, uh, the underclass.

Everyone wants to look like they are a member of a higher class than they actually are. But everyone also wants to avoid getting mistaken for a member of a poorer class. So for example, the middle-class wants to look upper-class, but also wants to make sure no one accidentally mistakes them for lower-class.

But there is a limit both to people's ambition and to their fear. No one has any hopes of getting mistaken for a class *two* levels higher than their own: a lower-class person may hope to appear middle-class, but their mannerisms, accent, appearance, peer group, and whatever make it permanently impossible for them to appear upper-class. Likewise, a member of the upper-class may worry about being mistaken for middle-class, but there is no way they will ever get mistaken for lower-class, let alone underclass.

So suppose we start off with a country in which everyone wears identical white togas. One day the upper-class is at one of their fancy upper-class parties, and one of them suggests that they all wear black togas instead, so everyone can recognize them and know that they're better than everyone else. This idea goes over well, and the upper class starts wearing black.

After a year, the middle class notices what's going on. They want to pass for upper-class, and they expect to be able to pull it off, so they start wearing black too. The lower- and underclasses have no hope of passing for upper-class, so they don't bother.

After two years, the lower-class notices the middle-class is mostly wearing black now, and they start wearing black to pass as middle-class. But the upper-class is very upset, because their gambit of wearing black to differentiate themselves from the middle-class has failed – both uppers and middles now wear identical black togas. So they conceive an ingenious plan to switch back to *white* togas. They don't worry about being confused with the white-

togaed underclass – no one could *ever* confuse an upper with a lower or under – but they will successfully differentiate themselves from the middles. Now the upper-class and underclass wear white, and the middle and lower classes wear black.

It's easy to see that this is the n = 4 version of the cellular automaton we just discussed.

Before I go on, an obvious objection – in a real world that doesn't work on "ticks", how do classes coordinate like this? Like, even if someone in the upper-class sent a super-secret message by butler to every single other member of the upper class saying "Tomorrow we all start wearing black, don't tell anyone else", within a day the rest of the world would notice, and the upper-class' advantage would be lost. And surely in our real world, where the upper-class has no way of distributing secret messages to every single cool person, this would be even harder. They'd have to announce their plan publicly, which would make the signal worthless.

There are some technical solutions to the problem. Upper class people are richer, and so can afford to about-face very quickly and buy an entirely new wardrobe. Upper class people have upper class friends, so it's easier for them to notice that black is 'in' and switch accordingly.

But I think the major solution is that there aren't only four classes, and no one is entirely sure what classes they can or can't pass for. The richest, trendiest person around wears something new, and either she is so hip that her friends immediately embrace it as a new trend, or she gets laughed at for going out in black when everyone *knows* all the cool people wear white. Her friends are either sufficiently hip that they then adopt the new trend and help it grow, or so unsure of themselves that they decide to stick with something safe, or so un-hip that when they adopt the new trend everyone laughs at them for being so clueless they think they can pull off being one of the cool people.

Or – you can't just copy someone else's outfit. That would be crass. So you have to understand the *spirit* of the fashion. But this is hard to get right if you're not familiar with it. The less exposure you have to the values and individuals who generated it, the more likely you'll get it wrong and end up looking like an idiot.

In other words, new trends carry social risk, and only people sufficiently clued-in and trendy can be sure the benefits outweigh the risks. But as the trend catches on, it becomes less risky, until eventually you see your Aunt Gladys wearing it because she saw something about it in a supermarket tabloid, and then all the hip people have to find a new trend.

There's another solution to this problem too: the upper class copies trends from the underclass. We saw this happen naturally on the 5th tick of the four-cell world, but it might be a more stable configuration than that model suggests. If the rich deliberately dress like the poor, then the middle-class have nowhere to go – if they try to ape the rich, they will probably just end up looking poor instead. It is only the rich, who are at no risk of ever being mistaken for the poor, who can pull this off.

Why do I like this model? It explains a lot of otherwise mysterious things about fashion.

Why does fashion change so darned often? Why can't people just figure out what's pretty, then stick to that?

Why is wearing last year's fashion such a faux pas? Shouldn't the response be "That person is wearing the second most fashionable outfit ever discovered; that's still pretty good"?

Why does fashion so often copy the outfits of the lower class (eg "ghetto chic"?) Why, if you are shopping for men's shirts, are there so many that *literally* say "GHETTO" on them in graffiti-like lettering?

And I don't think I'm a random nerd coming in here and telling fashion people that I understand them better than they understand

themselves. This seems to be how fashion people really think. Just look at the word "poser" (or possibly "poseur"). The thrust seems to be: "A person who is not of the group that is cool enough to wear this fashion is trying to wear this fashion! Get 'em!"

The big complication is that there is not one ladder of coolness going from "upper class" down to "underclass". There are businesspeople, intellectuals, punks, Goths – all of whom are trying to signal something different. And there's more than just white or black – hundreds of different colors, styles, and whatever.

But I think this is the fundamental generator that makes it all tick. In fact, I think this principle – counter-signaling hierarchies – is the fundamental generator that makes a lot of things tick.

II.

In the past two months I have inexplicably and very very suddenly become much more conservative.

This isn't the type of conservativism where I agree with any conservative *policies*, mind you. Those still seem totally wrong-headed to me. It's the sort of conservativism where, even though conservatives seem to be wrong about everything, often in horrible or hateful ways, they seem like probably mostly decent people deep down, whereas I have to physically restrain myself from going on Glenn Beck style rants about how much I hate leftists and how much they are ruining everything. Even though I mostly agree with the leftists whenever they say something.

(In fact, it seems like an important observation that there is a state of mind in which, no matter what your intelligence or rationality level, Glenn Beck or Rush Limbaugh-style rants against The Left seem justifiable and fun to listen to. I cannot communicate this state of mind and don't know why it occurs.)

At first I didn't notice this, because way back when I was a teenager and *very* leftist, I made a conscious decision that in order to counter my natural biases I should try to be as understanding

and friendly to conservatives as possible. I gradually got better and better at this and didn't notice that I was getting *too* good at it until it suddenly started to explode.

And now I am trying to figure out why that is.

Like all of you, my first thought was of course the pathogen stress theory of values. If conservative values are fueled by fear of contamination based on an inbuilt evolutionary reaction to the observed level of pathogen exposure, then my current work on an internal medicine hospital team – which is pretty heavy on the death and disease even for a doctor – would turn me super-conservative very quickly. But this hypothesis should mean that *all* doctors should be very conservative, which doesn't seem to be true. So scratch that.

Perhaps it's a natural effect of settling down, having a stable job, living in my own house, and being in a long-term relationship. But again, a lot of people seem to do all those things without becoming conservative. And none of that has changed in the past few months.

I do admit that, although I try to base my reasoned opinions on The Greater Good, a lot of my political emotions are based on fear, especially fear for my personal safety. I don't feel remotely threatened from the right – even when I meet anti-Semites who think all Jews should die, my feelings are mostly benevolent bemusement. I know if it ever came to any conflict between me and them, then short of them killing me instantly I would have *everyone in the world* on my side, and the possibility of it ending in any way other than with them in jail and me a hero who gets praised for his bravery in confronting them is practically zero. On the other hand, I feel massively threatened from the left, since the few times I got in a fight with *them* ended with me getting death threats and harrassment and feeling like everyone was on their side and I was totally alone. But nothing *new* of this sort has happened

in the past two months. That was probably a risk factor, but it can't have been the trigger.

I've been under a lot of stress lately – nothing serious, just very busy days at work with pretty much no free time (writing blog entries doesn't require free time. They just appear.) It wouldn't really surprise me if stress were related to conservativism. But I've been much more stressed in the past without this effect. Maybe work-related stress has some special ability to cause this effect? That would explain why so many working-class people with crappy jobs end up conservative.

The Left has been doing an unusual number of bad things in the past two months. I remember especially noticing the Eich incident and invasion of the Dartmouth administration building and related threats and demands. And then there was that thing with the national debate championships that is so horrible I still refuse to believe it and hold out hope against hope it turns out to be some absurdly irresponsible reporting or maybe a very very late April Fools' joke. But I feel like these sorts of things probably go on all the time, and my increased conservativism is the cause, and not the effect, of me noticing them. And I notice I don't feel the same level of cosmic horror when conservatives do something equally outrageous.

The explanation I like least is that it comes from reading too much neoreaction. I originally rejected this hypothesis because I don't believe most what I read. But I'm starting to worry that there are memes that, like Bohr's horseshoe, affect you whether you believe them or not: memes that crystallize the wrong pattern, or close the wrong feedback loop. I have long suspected social justice contains some of these. Now I worry neoreaction contains others.

In particular I worry about the neoreactionary assumption that leftism always increases with time, and that today's leftism confined to a few fringe idiots whom nobody really supports today becomes tomorrow's mainstream left and the day after tomorrow's

“you will be fired if you disagree with them”. Without me ever really evaluating its truth-value it has wormed its way into my brain and started haunting my nightmares.

Certain versions of it are certainly plausible. In 1960, only a handful of low status people were arguing that “sodomy laws” should be repealed, and they were all insisting that c’mon, obviously it would never go as far as gay *marriage*, we’re just saying you shouldn’t be put in jail for it. Meanwhile, fifty years later people are enforcing a rule that if you’re not on board with gay marriage, you shouldn’t be allowed to hold a high-status job.

Of course, many leftist views, even leftist social views, don’t spiral out of control like this. Support for abortion and gun control have stayed pretty stable for decades, radical feminism seems to have leveled off, and aside from global warming environmentalism has kind of faded into the background. But it’s impossible to predict which ones are going to spiral – to a 1960s conservative homosexuality would have seemed just about the *least* likely thing to catch on.

So now every time I read an article about horrible conservatives – like that South Carolina mayor – I can dismiss it as a couple of people doing dumb things and probably the system will take care of it. If it doesn’t take care of it by punishing him personally, it’ll take care of it by making people like him obsolete and judged poorly by posterity.

But every time I read an article about horrible leftists – like the one with the debate club – part of me freaks out and thinks – in twenty years, those are the people who are going to be getting me fired for disagreeing with them.

And every time I want to talk about it, I freak out and worry that soon they’ll start firing people for disagreeing with the idea that you should be able to fire people for disagreeing with ideas. Like, this could go *uncomfortably* far.

And so there is a dark and unpleasant Orwellian part of my brain that tells me: "If you want a vision of the future, imagine a hack misjudging a college debate – forever."

III.

But like I said, that's the explanation I like least. My favorite involves those cellular automata from before.

A friend recently pointed out that conservatives aren't, on average, very smart. He illustrated this with a graph of IQ vs. political belief which confirms that the left has a significant advantage.

But I look at my Facebook feed, and here is what I observe.

I see my high school classmates – a mostly unselected group of the general suburban California population – posting angry left stuff like "Ohmigod I just heard about that mayor in South Carolina WHAT A FUCKING BIGGOT!!!"

I see the people I think of as my intellectual equals posting things that are conspicuously nuanced – "Oh, I heard about that guy in South Carolina. Instead of knee-jerk condemnation, let's try to form some general principles out of it and see what it teaches us about civil society."

And I see the people I think of as the level above me posting extremely bizarre libertarian-conservative screeds making use of advanced mathematics that I can barely understand: "The left keeps saying that marriage as an institution isn't important. But actually, if we look at this from a game theoretic perspective, marriage and social trust and forager values are all in this complicated six-dimensional antifragile network, and it emergently coheres into a beneficial equilibrium if and only if the government doesn't try to shift the position of any of the nodes. Just as three eighteenth-century Frenchmen and a renegade Brazilian Marxist philosopher predicted. SO HOW COME THE IDIOTS ON THE LEFT KEEPS TRYING TO MAKE GOVERNMENT SHIFT THE POSITION OF THE NODES ALL THE TIME???!"

(I will proceed to describe this level extensionally: Jonathan Haidt, Bowling Alone, time discounting, public choice theory, the Hajnal line, contract law, Ross Douthat, incentives, polycentric anything, unschooling, exit rights)

And, I mean, I know the reason I get so many people trying to come up with bizarre mathematizations of politics is because those are the sorts of people I select as my friends. The part I don't get is why so many of them end up weird libertarian-conservative. Certainly not because I selected them for that. I don't even think they were weird libertarian-conservatives a few years ago when I met a lot of them. It just seems to have caught on.

And my theory is that in a world where the upper class wears black and the lower class wears white, they're the people who have noticed that the middle class is wearing black as well, and have decided to wear white to differentiate themselves.

It's the reverse of the 1950s. Assume you're a hip young intellectual in the 1950s. You see all these stodgy conservatives around you – I don't even know what "stodgy" means, I just know I'm legally obligated to use it to describe 1950s conservatives. You see Mrs. Grundy, chattering to her grundy friends about how *scandalous* it is that some people read books about sex, lecturing to the school board on how they had *better* enforce her values on the children or she will have some *very* harsh words to say to them.

And you think "Whatever else I am, I'm not going to be a mediocrity like Mrs. Grundy. I'm not going to *conform*." Which, in the 1950s, meant you became a leftist, and talked about how stodgy society was fundamentally oppressive, and how you were going to value *different* things, and *screw* what Mrs. Grundy thought.

And gradually this became sufficiently hip that even the slightly less hip intellectuals caught on and started making fun of Mrs. Grundy, and then people even less hip than that, until it became a

big pileup on poor Mrs. Grundy and anyone who wanted even the slightest claim to intellectual independence or personal integrity has to prove themselves by giving long dissertations on how terrible Mrs. Grundy is.

But when Mrs. Grundy herself joins the party, what then?

I mean, take that article on Dartmouth. A group of angry people, stopping just short of violence, invade a school building and make threats against the president unless he meets their demands. Every student must be forced to attend moral instruction classes inculcating their (the protesters') values. Offensive terms must be removed from the library. And the school must take care to admit people of the *right* race. When was the last time you could hear a story like that and have it be even *slightly* probably that the mob was rightist?

It's hard to argue that Mrs. Grundy is not a proud leftist by now, still chattering about how scandalous it is that people read books with the wrong values, still giving her terminally uncool speeches to the school board about how they had *better* enforce her values on the children (and if she can get the debate society on board as well, so much the better).

There must be overwhelming temptation among hip intellectuals to differentiate themselves from Mrs. Grundy by shifting rightward.

And perhaps so far this has been kept in check by the second rule of our cellular automaton – you can't take a position that would get you plausibly confused for a person of lower class than you.

I was tickled by a conversation between two doctors I recently heard in a hospital hallway:

> **Doctor 1:** My daughter just got a full scholarship into a really good university in Georgia.
> **Doctor 2:** Congratulations!
> **Doctor 1:** Thanks! But I'm hoping she'll choose somewhere closer to home.

Doctor 2: Why? Because you want to be able to visit her more?
Doctor 1: There's that. But the other problem is that the South is full of *those people*.
Doctor 2: So? Colleges are like their own world. Your daughter probably won't even encounter many of them.
Doctor 1: I know. But I keep worrying that just by being there, she'll make friends with them, and then end up bringing one home as a boyfriend.

"Those people" is my replacement, not the original term used by the doctor involved. The doctor involved said a much less polite word.

She said "fundies".

Fundies – in all of their Bible-beating gun-owning cousin-marrying stereotypicalness – have so far served as the Lower Class With Which One Must Not Allow One's Self To Be Confused. But I think that's changing. Sorting mechanisms are starting to work so well that, at the top, the fundies just aren't plausible. In our model, people from class N can be confused with class N-1, but never with class N-2. But as the barber-pole movement of fashion creeps downward, fundies are starting to become two classes below certain people at the top, and those people no longer risk misidentification.

I notice that, no matter how many long rants against feminism I write, everyone continues to assume I am a feminist. It's like, "He doesn't make too many spelling errors, his writing isn't peppered with racial slurs – he's got to be a feminist. He probably just forgot the word 'not' in each of his last 228 sentences."

And I wonder if maybe the reason why I am outraged by the debate team but not by the South Carolina mayor isn't that I feel a greater threat from the debate team, but because I feel like there is a greater threat of me being *mistaken* for the debate team. If

impotent expressions of outrage divorced from any effort to change things are ways of saying "I'm not like this! I promise!" And I get less outraged than some other people about South Carolina because I feel confident enough in my intelligence that I don't worry anyone will mistake me for a fundie. But I feel less confident no one could mistake me for the sort of person who judged those debate championships, so I need to shout at them to show I'm Not Like That. This would actually explain a lot.

If some intellectuals no longer need to worry about being mistaken for fundies, that frees them to finally breath a sigh of relief and start making fun of Mrs. Grundy again. And that means they've got to become conservatives, or libertarians, or anything, anything at all, except for leftists.

So far it is just a few early adopters – the intellectual equivalent of the very trendy people who start wearing some outrageous fashion and no one knows if it is going to catch on or whether they will be soundly mocked for it.

And they are having a really difficult time, because a lot of conservative ideas aren't that great. Like, reality leaves you a lot of degrees of freedom when you're deciding your political self-presentation, but it doesn't leave you an *infinite* number of degrees of freedom, and the project of creating something that is both anti-leftist enough to serve as a fashion statement but reality-based enough not to be dumb is still going on. The reactionaries are doing an excellent job maximizing the "anti-leftist" criterion. The "reality-based" criterion is a harder egg to crack, but it makes me think of Drew Summitt, Athrelon, and some of SarahC's more political moments.

As the Commissioner puts it, "Evolution is at work here, but just what is evolving remains to be seen."

When I put it like this, I realize I'm not becoming more conservative at all. I'm becoming anti-leftist. Actually, put that

way a lot of people seem to be anti-leftist. I can't think of a single specific policy proposal supported by Glenn Beck. Can you?

And I think the best explanation is that all my hip friends who I want to be like are starting to be conservative or weird-libertarian or some variety of non-leftist, and Mrs. Grundy is starting to become very obviously leftist and getting grundier by the day, and so the fashion-conscious part of my brain, the much-abused and rarely-heeded part that tells me "No, you can't go to work in sweatpants, even though it would be much more comfortable", is telling me "QUICK, DISENGAGE FROM UNCOOL PEOPLE AND START ACTING LIKE COOL PEOPLE RIGHT NOW."

And I said this is my favorite of all the explanations. Why?

Because if it's true, and it spreads beyond a couple of little subcultures, it means my worst fears are misplaced. The future isn't a foot stamping on the face of a a college debate team forever. It's people – or at least some people – rolling their eyes at those people and making fake vomiting noises. And then going too far, until other people have to roll their eyes at *those* people. And so on. Instead of a death spiral we get a pendulum, swinging back and forth.

But I would hope for something even better than that. Like, at each swing of the pendulum, people learn a little. I was really impressed with how many smart and decent people thought that the Eich thing was wrong (…and wore kilts, and played bagpipes…shut up). Fashion does not accrete, but maybe reality does. And I would like to think that the rationalist movement is a part of that. And if that's true, that's a way in which reality will eventually come to overpower fashion and the arc of the universe might tend toward justice after all.

Weak Men are Superweapons

I.

There was an argument on Tumblr which, like so many arguments on Tumblr, was terrible. I will rephrase it just a little to make a point.

Alice said something along the lines of "I hate people who frivolously diagnose themselves with autism without knowing anything about the disorder. They should stop thinking they're 'so speshul' and go see a competent doctor."

Beth answered something along the lines of "I diagnosed myself with autism, but only after a lot of careful research. I don't have the opportunity to go see a doctor. I think what you're saying is overly strict and hurtful to many people with autism."

Alice then proceeded to tell Beth she disagreed, in that special way only Tumblr users can. I believe the word "cunt" was used.

I notice two things about the exchange.

First, why did Beth take the bait? Alice said she hated people who *frivolously* self-diagnosed *without knowing anything about the disorder*. Beth clearly was not such a person. Why didn't she just say "Yes, please continue hating these hypothetical bad people who are not me"?

Second, why did *Alice* take the bait? Why didn't she just say "I think you'll find I wasn't talking about you?"

II.

One of the cutting-edge advances in fallacy-ology has been the weak man, a terribly-named cousin of the straw man. The

straw man is a terrible argument nobody really holds, which was only invented so your side had something easy to defeat. The weak man is a terrible argument that only a few unrepresentative people hold, which was only *brought to prominence* so your side had something easy to defeat.

For example, “I am a proud atheist and I don’t like religion. Think of the terrible things done by religion, like the actions of the Westboro Baptist Church. They try to disturb the funerals of heroes because they think God hates everybody. But this is horrible. Religious people can’t justify why they do things like this. That’s why I’m proud to be an atheist.”

It’s not a straw man. There really is a Westboro Baptist Church, for some reason. But one still feels like the atheist is making things just a little too easy on himself.

Maybe the problem is that the atheist is indirectly suggesting that Westboro Baptist Church is typical of religion? An implied falsehood?

Then suppose the atheist posts on Tumblr: “I hate religious people who are rabidly certain that the world was created in seven days or that all their enemies will burn in Hell, and try to justify it through ‘faith’. You know, the sort of people who think that the Bible has all the answers and who hate anyone who tries to think for themselves.”

Now there’s practically no implication that these people are typical. So that’s fine, right?

On the other side of the world, a religious person is writing “I hate atheists who think morality is relative, and that this gives them the right to murder however many people stand between them and a world where no one is allowed to believe in God”.

Again, not a straw man. The Soviet Union contained several million of these people. But if you're an atheist, would you just let this pass?

How about "I hate black thugs who rob people"?

What are the chances a black guy reads that and says "Well, good thing I'm not a thug who robs people, he'll probably *love* me"?

III.

What is the problem with statements like this?

First, they are meant to re-center a category. Remember, people think in terms of categories with central and noncentral members – a sparrow is a central bird, an ostrich a noncentral one. But if you live on the Ostrich World, which is inhabited only by ostriches, emus, and cassowaries, then probably an ostrich seems like a pretty central example of 'bird' and the first sparrow you see will be fantastically strange.

Right now most people's central examples of religion are probably things like your local neighborhood church. If you're American, it's probably a bland Protestant denomination like the Episcopalians or something.

The guy whose central examples of religion are Pope Francis and the Dalai Lama is probably going to have a different perception of religion than the guy whose central examples are Torquemada and Fred Phelps. If you convert someone from the first kind of person to the second kind of person, you've gone most of the way to making them an atheist.

More important, if you convert a culture from thinking in the first type of way to thinking in the second type of way, then religious people will be unpopular and anyone trying to make a religious argument will have to spend the first five minutes

of their speech explaining how they're not Fred Phelps, honest, and no, they don't picket any funerals. After all that time spent apologizing and defending themselves and distancing themselves from other religious people, they're not likely to be able to make a very rousing argument for religion.

IV.

In Cowpox of Doubt, I mention the inoculation effect. When people see a terrible argument for an idea get defeated, they are more likely to doubt the idea later on, even if much better arguments show up.

Put this in the context of people attacking the Westboro Baptist Church. You see the attacker win a big victory over "religion", broadly defined. Now you are less likely to believe in religion when a much more convincing one comes along.

I see the same thing in atheists' odd fascination with creationism. Most of the religious people one encounters are not young-earth creationists. But these people have a dramatic hold on the atheist imagination.

And I think: well, maybe if people see atheists defeating a terrible argument for religion enough, atheists don't *have to* defeat any of the others. People have already been inoculated against religion. "Oh, yeah, that was the thing with the creationism. Doesn't seem very smart."

If this is true, it means that all religious people, like it or not, are in the same boat. An atheist attacking creationism becomes a deadly threat for the average Christian, even if that Christian does not herself believe in creationism.

Likewise, when a religious person attacks atheists who are moral relativists, or communists, or murderers, then all atheists

have to band together to stop it somehow or they will have successfully poisoned people against atheism.

V.

This is starting to sound a lot like something I wrote on my old blog about superweapons.

I suggested imagining yourself in the shoes of a Jew in czarist Russia. The big news story is about a Jewish man who killed a Christian child. As far as you can tell the story is true. It's just disappointing that everyone who tells it is describing it as "A Jew killed a Christian kid today". You don't want to make a big deal over this, because no one is saying anything objectionable like "And so all Jews are evil". Besides you'd hate to inject identity politics into this obvious tragedy. It just sort of makes you uncomfortable.

The next day you hear that the local priest is giving a sermon on how the Jews killed Christ. This statement seems historically plausible, and it's part of the Christian religion, and no one is implying it says anything about the Jews today. You'd hate to be the guy who barges in and tries to tell the Christians what Biblical facts they can and can't include in their sermons just because they offend you. It would make you an annoying busybody. So again you just get uncomfortable.

The next day you hear people complain about the greedy Jewish bankers who are ruining the world economy. And really a disproportionate number of bankers are Jewish, and bankers really do seem to be the source of a lot of economic problems. It seems kind of pedantic to interrupt every conversation with "But also some bankers are Christian, or Muslim, and even though a disproportionate number of bankers are Jewish that doesn't mean the Jewish bankers are

disproportionately active in ruining the world economy compared to their numbers." So again you stay uncomfortable.

Then the next day you hear people complain about Israeli atrocities in Palestine (what, you thought this was past czarist Russia? This is future czarist Russia, after Putin finally gets the guts to crown himself). You understand that the Israelis really do commit some terrible acts. On the other hand, when people start talking about "Jewish atrocities" and "the need to protect Gentiles from Jewish rapacity" and "laws to stop all this horrible stuff the Jews are doing", you just feel worried, even though you personally are not doing any horrible stuff and maybe they even have good reasons for phrasing it that way.

Then the next day you get in a business dispute with your neighbor. Maybe you loaned him some money and he doesn't feel like paying you back. He tells you you'd better just give up, admit he is in the right, and apologize to him – because if the conflict escalated everyone would take his side because he is a Christian and you are a Jew. And everyone knows that Jews victimize Christians and are basically child-murdering Christ-killing economy-ruining atrocity-committing scum.

You have been boxed in by a serious of individually harmless but collectively dangerous statements. None of them individually referred to you – you weren't murdering children or killing Christ or owning a bank. But they ended up getting you in the end anyway.

Depending on how likely you think this is, this kind of forces Jews together, makes them become strange bedfellows. You might not like what the Jews in Israel are doing in Palestine. But if you think someone's trying to build a superweapon against you, and you don't think you can differentiate yourself

from the Israelis reliably, it's in your best interest to defend them anyway.

VI.

I wrote the superweapon post to address some of my worries about feminism, so it would not be surprising at all if we found this dynamic there.

Feminists tend to talk about things like "Men tend to silence women and not respect their opinions" or "Men treat women like objects rather than people" or "Men keep sexually harassing women even when they make it clear they're not interested".

Put like that, it's obvious why men might complain. But maybe some of the more sophisticated feminists say "Some men tend to silence women and not respect their opinions". Or "Some men keep sexually harassing women even when they make it clear they're not interested."'

And the weak-man-superweapon model would suggest that even this weakened version would make lots of men really uncomfortable.

From feminist website Bitchtopia (look, I don't name these websites, I just link to them): Not All Men Are Like That:

> I've heard this counter-argument almost every single time I've tried to bring up a feminist issue with a man: "but not all men are like that!"…
>
> Having to point out that not every man exhibits explicitly harmful behavior allows for oppression to continue because having to say "some men do harmful things" gives oppressors peace of mind…

> Sure, white men–you were brought up to feel entitled to anything you wanted and now you see anyone trying to have opportunities equal to yours as a threat…
>
> When you say, "not all men are like that!" what you're really saying is, "I don't want to have to think about my privilege as a white man, so I'm going to try to defer the blame to other guys because I clearly don't act like that."
>
> Nice try.

Remember, not wanting to be stereotyped based solely on your sex is the *most* sexist thing!

This is not just an idiosyncracy of Bitchtopia (look! I'm sorry! I swear I didn't name that website!). There's also an entire notallmenarelikethat dot tumblr dot com (of *course* there is) and it's now a feminist meme abbreviated NAMALT.

But of course, it's not just feminists. The gender-flipped version of feminism has the same thing. From men's rights blog "The Spearhead", which is not quite as badly named but still kind of funny if you think of it in a Freudian way:

> Talking about the current sad state of dating and marriage in the USA will often elicit "Not All Women Are Like That" or NAWALT.
>
> The first thing is not to contradict whoever makes that claim. Why? Because it is true. Not all women are skanks, attention whores or predators. The MRA cause is not helped by attacking people who speak truthfully.
>
> [But the consequence of a] false positive is that a man ends up married to a skank, sociopath or gold digger. The cost of bad wife selection is so high that he is forced to

> turn away good women for fear of mistakenly choosing a bad one.

More polite and scientific than the feminist version, but the point is he expects men's rights readers to be so familiar with "not all women are like that" that he's perfectly comfortably abbreviating it NAWALT. Apparently there's even a NAWALT video.

I don't know where to find neo-Nazi blogs, but I'll bet if there are some, they have places where they talk about how annoying it is when people try to distract from the real issues by using the old NAJALT.

VII.

But I shouldn't make fun of NAJALT. There really are two equal and opposite problems going on here.

Imagine you're an atheist. And you keep getting harassed by the Westboro Baptist Church. Maybe you're gay. Maybe you're not. Who knows why they do what they do? Anyway, they throw bricks through your window and send you threatening letters and picket some of your friends' funerals.

And you say "People! We really need to do something about this Westboro Baptist Church! They're horrible people!"

And you are met by a wall of religious people saying "Please stop talking about the Westboro Baptist Church, you are making us look really bad and it's unfair because not all religious people are like that."

And you say "I really am not that interested in religion, I just want them to stop throwing bricks through my window."

And they say "Hey! I thought we told you to stop talking about them! You are unfairly discrediting us through the

inoculation effect! That is epistemically unvirtuous!"

So the one problem is that people have a right not to have unfair below-the-belt tactics used to discredit them without ever responding to their real arguments.

And the other problem is that victims of nonrepresentative members of a group have the right to complain, even though those complaints will unfairly rebound upon the other members of that group.

Atheists who talk about the Westboro Baptist Church may be genuinely concerned about the Westboro Baptist Church. Or they may be unfairly trying to tar all religious people with that brush. Religious people have to fight back, even though the Westboro Baptists don't deserve their support, because otherwise the atheists will have a superweapon against them. Thus, a stupid fight between atheists who don't care about Westboro and religious people who don't support them.

VIII.

This gives me some new views on political coalitions. I always thought that having things like political parties was stupid. Instead of identifying as a liberal and getting upset when someone insulted liberals or happy when someone praised liberals, I should say "These are my beliefs. There are other people who believe approximately the same thing, but the differences are sufficient that I just want to be judged on my own individual beliefs alone."

The problem is, that doesn't work. It's not my decision whether or not I get to identify with other liberals or not. If other people think of me as a liberal, then anything other liberals do is going to reflect, positively or negatively, on me. And I'm going to have to join in the fight to keep liberals from being completely discredited, or else the fact that I didn't share

any of the opinions they were discredited for isn't going to save me. I will be Worst Argument In The World-ed and swiftly dispatched.

In the example we started with, Beth chose to stand up for the people who self-diagnosed autism without careful research. This wasn't because she considered herself a member of that category. It was because she decided that self-diagnosed autistics were going to stand or fall as a group, and if Alice succeeded in pushing her "We should dislike careless self-diagnosees" angle, then the fact that she wasn't careless wouldn't save her.

Alice, for her part, didn't bother bringing up that she never accused Beth of being careless, or that Beth had no stake in the matter. She saw no point in pretending that boxing in Beth and the other careful self-diagnosers in with the careless ones wasn't her strategy all along.

You Kant Dismiss Universalizability

I.

Like most right-thinking people, I'd always found Immanuel Kant kind of silly. He was the standard-bearer for naive deontology, the "rules are rules, so follow them even if they ruin everything" of moral philosophy.

But lately, I've been starting to pick up a different view. There may have been some subtleties I was missing, almost as if one of the most universally revered thinkers of the western philosophical tradition wasn't a total moron.

I was delighted to see nydwracu say something similar in the comments to my recent post:

> I [now] realize that Kant is not actually completely ridiculous like I once thought he was

I don't know if it's just that nydwracu and I have been thinking about some of the same problems lately, but he took the words right out of my mouth.

I'm not a Kant scholar. I'm not qualified to explain what Kant thought, and it's possible the arguments I express as Kantian here are going to be arguments of a totally different person who merely reminds me of Kant in some ways. James Donald's objections to steelmanning are well taken, so I will not call this a steel man of a guy who is too dead to correct me if I am wrong. At best I will call this post Kant-aligned.

First, I want to take another look at one of Kant's most-reviled arguments: that you should truthfully tell a murderer who wants to kill your friend where she is hiding.

Second, I want to talk about how I find myself using Kantian principles in my own morality.

And third, I want to talk about big unanswered questions and the reason this still isn't technical enough for me to be comfortable with.

II.

Kant gives the following dilemma. Suppose that an axe murderer comes to your door and demands you tell him where your friend is, so that he can kill her. Your friend in fact is in your basement. You lie and tell the murderer your friend is in the next town over. He heads off to the next town, and while he's gone you call the police and bring your friend to safety.

Most people would say that the lie is justified. Kant says it isn't, because lying.

I think most people understand his argument as follows: you think "I should lie". But suppose everyone thought that all the time. Then everyone would lie to everyone else, and that would be horrible.

But Kant's categorical imperative doesn't urge us to reject actions which, if universalized, would be horrible. That's rule utilitarianism, sort of. Kant urges us to reject actions which, if universalized, would be self-defeating or contradictory.

Suppose it was everyone's policy to lie to axe murderers who asked them where their friends were. Well, then axe murderers wouldn't even bother asking.

Which doesn't sound like a sufficiently terrible dystopia to move us very much. So let me reframe Kant's example.

Suppose you are a prisoner of war. Your captors tell you they want to kill your general, a brilliant leader who has led your side to victory after victory. They have two options. First, a

surgical strike against her secret headquarters, killing her and no one else. Second, nuking your capital city. They would prefer to do the first, because they're not monsters. But if they have to nuke your capital, they'll nuke your capital. So they show you a map of your capital city and say "Please point out your general's headquarters and we'll surgical-strike it. But if you don't, we'll nuke the whole city."

You decide to lie. You point to a warehouse you know to be abandoned. Your captors send a cruise missile that blows up the warehouse, killing nobody. Then they hold a huge party to celebrate the death of the general. Meanwhile, the real general realizes she's in danger and flees to an underground shelter. With her brilliant tactics, your side wins the war and you are eventually rescued.

So what about now? Was your lie ethical?

Kant would point out that if it was known to be everyone's policy to lie about generals' locations, your captors wouldn't even ask. They'd just nuke the city, killing everyone.

Your captors are offering you a positive-sum bargain: "Normally, we would nuke your capital. But you don't want that and we don't want that. So let's make a deal where you tell us where your general is and we only kill that one person. That leaves both of us better off."

If it is known to everyone that prisoners of war always lie in this situation, it would be impossible to offer the positive-sum bargain, and your enemies would resort to nuking the whole city, which is worse for both of you.

So when Kant says not to act on maxims that would be self-defeating if universalized, what he means is "Don't do things that undermine the possibility to offer positive-sum bargains."

This is *very* reminiscent of Parfit's Hitchhiker. Remember that one? You are lost in the desert, about to die. A very selfish man drives by in his dune buggy, sees you, and offers to take you back to civilization for $100. You don't have any money on you, but you promise to pay him $100 once you're back to civilization and its many ATMs. The very selfish man agrees and drives you to safety. Once you're safe, you say "See you later, sucker!" and run off.

The selfish man's "I'll bring you back to civilization for $100" offer is a positive-sum bargain. You would rather lose $100 than die. He would rather gain $100 and lose a few hours bringing you to the city than continue on his way. So you both gain.

But if everyone were omniscient and knew that people who promise $100 will never really pay, or if your decision not to pay could somehow affect his willingness to make you the offer in the first place, the ability to make the positive-sum bargain disappears.

On this model, Kant isn't being a weird super-anal stickler for meaningless rules at all. He's being the most practical person around: don't do things that spoil people's ability to make a profit.

(and sort of pre-inventing decision theory)

(man, it's a good thing everyone is omniscient and the future can cause the past, or else we'd never be able to ground morality *at all*)

III.

A while back I suggested it is wrong to fire someone for being anti-gay, because if every boss said "I will fire my employees whom I disagree with politically", or every mob of angry

people said “We will boycott companies until they fire the people we disagree with politically” then no one who’s not independently wealthy could express any political opinions or dare challenge the status quo, and the world would be a much sadder place.

This is not strictly Kantian. “The world would be a much sadder place” is not self-defeating or a contradiction.

But it could still be framed as a positive-sum bargain. In a world where all the leftists refused to hire rightists, and all the rightists refused to hire leftists, everything would be about the same except that everyone’s job opportunities would be cut in half. If the people in such a world were halfway rational, they would make a deal that rightists agree to hire leftists if leftists agree to hire rightists. This would clearly be positive-sum.

This is easy to say in natural language like this. But when you try to make it more formal it gets really sketchy real quick.

Let’s say Paula the Policewoman is arresting Robby the Robber (she caught him by noticing his name was Robby in a world where everyone’s name sounds like their most salient characteristic). No doubt she thinks she is following the maxim “Police officers should arrest robbers”. But what about other maxims that lead to the same action?

1. Police officers should arrest people
2. Everyone should arrest robbers
3. Paula should arrest Robby
4. Paula should arrest other people
5. Everyone should arrest Robby
6. Everyone should arrest EVERYONE ELSE IN THE WORLD

This sounds kind of silly in this context, but in more complicated situations the entire point hinges upon it.

Levi the Leftist, who owns a restaurant called Levi's Lentils, finds out that his head waiter, Riley the Rightist, is a homophobe (in Levi's defense, he thought he was safe to hire him because his name wasn't Homer). He fires Riley, who ends out on the street.

Candice the Kantian condemns him, saying "What if that were to become a general rule? Then nothing would change except everyone only has half as many job opportunities."

Levi says "Oh, I see your problem. You think my maxim is 'fire people with different politics than me'. But that's not my maxim at all. My maxim is 'fire people who are homophobic'. If that becomes universalized, it will be a great victory for gay people everywhere, but no one whose politics I agree with will suffer at all."

In fact, Levi might claim his maxim is any one of the following:

1. Everyone should fire people they disagree with politically
2. Everyone should fire people who are politically on the right
3. Everyone should fire people who discriminate against minority groups
4. Everyone should fire people who are homophobic
5. Everyone should fire people who are mean and hateful
6. Everyone should fire people who hold positions that are totally beyond the pale and can't possibly be supported rationally

(before I get yelled at in the comment section, I'm not necessarily claiming all these maxims accurately describe Riley, just that Levi might think they do)

(5) runs into this problem where you can never say "fire people who are mean and hateful" without it *in fact* meaning "fire people whom *you think* are mean and hateful".

Presumably all the rightist bosses will find good reasons to think their leftist employees are mean and hateful.

There seems to be some sense in which we also want to protest (2), say that if Levi is allowed to use (2), then that instantly morphs to rightist bosses being allowed to say "everyone should fire people who are politically on the left". But just saying "universalizability!" doesn't automatically let us do that.

(3) seems even sneakier. It is in fact the maxim promoted by the people who are actually doing the firing, since they seem to have some inkling that universalizability and "fairness" are important. And it sounds totally value-neutral and universalizable. And yet I feel like if we allow Levi to say this, then some rightist will say actually *his* maxim is "everyone should fire people who want to undermine traditional cultural institutions", and the end result will be the same old "job opportunities halved for everyone".

IV.

This is a hard problem. The best solution I can think of right now is to go up a meta-level, to say "universalize as if the process you use to universalize would itself become universal".

Suppose I am very greedy, and I lie and steal and cheat to get money. I say "Well, my principle is to always do whatever gets Scott the most money". This soooooooorta checks out. If it were universalized – and everyone acted on the principle "Always do whatever gets Scott the most money", well, I wouldn't mind that at all.

But if we say "universalize as if the process you use to universalize would itself become universal", then we assume that if I try to universalize to "do what gets Scott the most

money", then Paula will universalize to "do what gets Paula the most money" and Levi will universalize to "do what gets Levi the most money" and we'll all be lying and cheating and stealing from one another and no one will be very happy at all.

(Kant notes that this also satisfies his original, stricter "self-defeating contradiction" criterion. If we all try to steal from each other, then private property becomes impossible, the economy collapses, and the stuff we want isn't there to steal. I don't know if I like this; it seems a little forced. But even if contradictoriness is forced, badness seems incontravertible)

As for Levi, he knows that if he universalizes to "everyone should fire people who discriminate against minority groups", his process is "pick out a political value that's important to me and excludes a lot of potential employees, then say everyone should fire people who disagree with it". This is sufficient to assume rightists will do the same and we'll be back at half-as-many-jobs.

Next problem. Suppose I am a very rich and very selfish tycoon. I say "No one should worry about helping the needy". I am perfectly happy with this being universalized, because it saves me from having to waste my time helping the needy. Although other people also won't help the needy, I'm a super-rich tycoon and that's no skin off *my* back.

We can climb part of the way out of this pit with meta-universalizability. We say "If I say things like this, everyone will only act on maxims that benefit them personally and appeal to their own idiosyncratic characteristics, rather than the ones that most benefit everyone."

But I worry that this isn't enough. Suppose I'm not just a tycoon, I'm a super-rich and powerful tyrannical king. I come up with maxims like "Everyone do what the tyrant says or be

killed!" Candice the Kantian warns "If you do that, everyone will come up with maxims that benefit them personally, and the moral law will be weakened."

And so I kill Candice for disagreeing with me.

If you are so much stronger than other people that you are immune to their counter-threats, you can get away with doing pretty much anything under this perversion of not-at-all-like-Kant we've wandered into.

We might have gotten so far from Kant at this point that we've stumbled into Rawls. Put up a veil of ignorance and the problem vanishes.

V.

What about utiltarianism?

I would love to universalize the maxim "Do whatever most increases Scott's utility".

Given concerns of meta-univeralizability above, I might end up instead wanting to universalize "Do whatever most increases global utility".

This seems certain, maybe even provable, if you throw in the veil of ignorance accessory.

Utilitarianism has a lot of the same problems universalizability does. A very stupid utilitarian would automatically condemn Levi for firing Riley since now Riley is unemployed and this lowers his utility. More sophisticated utilitarians would have to take into account the various society-wide effects of Levi setting a precedent here. I think that's what Mill's rule utilitarianism tries to do and what precedent utilitarianism tries to do as well. The problem is that it's really hard to figure out what rules and precedents have how much weight. Universalizability kind of plows through some of those

objections like a giant steamroller. It probably prevents a couple of little incidents where you could steal something or kill someone to gain a little extra utility, but it more than makes up for it in vastly increasing social trust and ability for positive-sum deals.

I'm not sure whether consequentialism is prior to universalizability ("universalize maxims because if you don't you'll end up losing out on possible positive-sum games and cutting your job offers in half"), whether universalizability is prior to consequentialism ("be a consequentialist, because that is a maxim everyone could agree on"), or whether they're like a weird ouroboros constantly eating itself.

I think maybe the idea I like best is that consequentialism is prior to universalizability is prior to any particular version of utilitarianism.

Because if universalizability is prior, that would be an interesting way to explore some of the problems with utilitarianism. For example, should we count pleasure or preferences? I don't know. Let's see what everyone would agree on.

Does everyone have to donate all of their money to the most efficient charity all the time? Well, if you were behind the veil of ignorance helping frame the moral law, would you put that in?

Does everyone have to prefer torture to dust specks? You're behind the veil of ignorance, you don't know if you'll be a dust speck person or a torture person, what do *you* think?

I think this is a good point to remember the blog tagline and admit I am still confused, but on a higher level and about more important things.

I Can Tolerate Anything Except the Outgroup

[Content warning: Politics, religion, social justice, spoilers for "The Secret of Father Brown". This isn't especially original to me and I don't claim anything more than to be explaining and rewording things I have heard from a bunch of other people. Unapologetically America-centric because I'm not informed enough to make it otherwise. Try to keep this off Reddit and other similar sorts of things.]

I.

In Chesterton's *The Secret of Father Brown*, a beloved nobleman who murdered his good-for-nothing brother in a duel thirty years ago returns to his hometown wracked by guilt. All the townspeople want to forgive him immediately, and they mock the titular priest for only being willing to give a measured forgiveness conditional on penance and self-reflection. They lecture the priest on the virtues of charity and compassion.

Later, it comes out that the beloved nobleman did *not* in fact kill his good-for-nothing brother. The good-for-nothing brother killed the beloved nobleman (and stole his identity). *Now* the townspeople want to see him lynched or burned alive, and it is only the priest who – consistently – offers a measured forgiveness conditional on penance and self-reflection.

The priest tells them:

> It seems to me that you only pardon the sins that you don't really think sinful. You only forgive criminals when they commit what you don't regard as crimes, but rather as conventions. You forgive a conventional duel just as you forgive a conventional divorce. You forgive because there isn't anything to be forgiven.

He further notes that this is why the townspeople can self-righteously consider themselves more compassionate and forgiving than he is. Actual forgiveness, the kind the priest

needs to cultivate to forgive evildoers, is really really hard. The fake forgiveness the townspeople use to forgive the people they like is really easy, so they get to boast not only of their forgiving nature, but of how much nicer they are than those mean old priests who find forgiveness difficult and want penance along with it.

After some thought I agree with Chesterton's point. There are a lot of people who say "I forgive you" when they mean "No harm done", and a lot of people who say "That was unforgiveable" when they mean "That was genuinely really bad". Whether or not forgiveness is *right* is a complicated topic I do not want to get in here. But since forgiveness is generally considered a virtue, and one that many want credit for having, I think it's fair to say you only earn the right to call yourself 'forgiving' if you forgive things that genuinely hurt you.

To borrow Chesterton's example, if you think divorce is a-ok, then you don't get to "forgive" people their divorces, you merely ignore them. Someone who thinks divorce is abhorrent can "forgive" divorce. *You* can forgive theft, or murder, or tax evasion, or something *you* find abhorrent.

I mean, from a utilitarian point of view, you are still doing the correct action of not giving people grief because they're a divorcee. You can have all the Utility Points you want. All I'm saying is that if you "forgive" something you don't care about, you don't earn any Virtue Points.

(by way of illustration: a billionaire who gives $100 to charity gets as many Utility Points as an impoverished pensioner who donates the same amount, but the latter gets a lot more Virtue Points)

Tolerance is *definitely* considered a virtue, but it suffers the same sort of dimished expectations forgiveness does.

The Emperor summons before him Bodhidharma and asks: "Master, I have been tolerant of innumerable gays, lesbians, bisexuals, asexuals, blacks, Hispanics, Asians, transgender people, and Jews. How many Tolerance Points have I earned for my meritorious deeds?"

Bodhidharma answers: "None at all".

The Emperor, somewhat put out, demands to know why not.

Bodhidharma asks: "Well, what do you think of gay people?"

The Emperor answers: "What do you think I am, some kind of homophobic bigot? Of course I have nothing against gay people!"

And Bodhidharma answers: "Thus do you gain no merit by tolerating them!"

II.

If I had to define "tolerance" it would be something like "respect and kindness toward members of an outgroup".

And today we have an almost unprecedented situation.

We have a lot of people – like the Emperor – boasting of being able to tolerate everyone from every outgroup they can imagine, loving the outgroup, writing long paeans to how great the outgroup is, staying up at night fretting that somebody else might not like the outgroup enough.

And we have those same people absolutely *ripping* into their in-groups – straight, white, male, hetero, cis, American, whatever – talking day in and day out to anyone who will listen about how terrible their in-group is, how it is responsible

for all evils, how something needs to be done about it, how they're ashamed to be associated with it at all.

This is really surprising. It's a total reversal of everything we know about human psychology up to this point. No one did any genetic engineering. No one passed out weird glowing pills in the public schools. And yet suddenly we get an entire group of people who conspicuously love their outgroups, the outer the better, and gain status by talking about how terrible their own groups are.

What is going on here?

III.

Let's start by asking what exactly an outgroup is.

There's a very boring sense in which, assuming the Emperor's straight, gays are part of his "outgroup" ie a group that he is not a member of. But if the Emperor has curly hair, are straight-haired people part of his outgroup? If the Emperor's name starts with the letter 'A', are people whose names start with the letter 'B' part of his outgroup?

Nah. I would differentiate between multiple different meanings of outgroup, where one is "a group you are not a part of" and the other is…something stronger.

I want to avoid a very easy trap, which is saying that outgroups are about how different you are, or how hostile you are. I don't think that's quite right.

Compare the Nazis to the German Jews and to the Japanese. The Nazis were very similar to the German Jews: they looked the same, spoke the same language, came from a similar culture. The Nazis were totally different from the Japanese: different race, different language, vast cultural gap. But although one could *imagine* certain situations in which the

Nazis treated the Japanese as an outgroup, in practice they got along pretty well. Heck, the Nazis were actually moderately friendly with the Chinese, even when they were technically at war. Meanwhile, the conflict between the Nazis and the German Jews – some of whom didn't even realize they were anything other than German until they checked their grandparents' birth certificate – is the stuff of history and nightmares. Any theory of outgroupishness that naively assumes the Nazis' natural outgroup is Japanese or Chinese people will be totally inadequate.

And this isn't a weird exception. Freud spoke of the narcissism of small differences, saying that "it is precisely communities with adjoining territories, and related to each other in other ways as well, who are engaged in constant feuds and ridiculing each other". Nazis and German Jews. Northern Irish Protestants and Northern Irish Catholics. Hutus and Tutsis. South African whites and South African blacks. Israeli Jews and Israeli Arabs. Anyone in the former Yugoslavia and anyone else in the former Yugoslavia.

So what makes an outgroup? Proximity plus small differences. If you want to know who someone in former Yugoslavia hates, don't look at the Indonesians or the Zulus or the Tibetans or anyone else distant and exotic. Find the Yugoslavian ethnicity that lives closely intermingled with them and is most conspicuously similar to them, and chances are you'll find the one who they have eight hundred years of seething hatred toward.

What makes an unexpected in-group? The answer with Germans and Japanese is obvious – a strategic alliance. In fact, the World Wars forged a lot of unexpected temporary pseudo-friendships. A recent article from War Nerd points out that the British, after spending centuries subjugating and despising the

Irish and Sikhs, suddenly needed Irish and Sikh soldiers for World Wars I and II respectively. “Crush them beneath our boots” quickly changed to fawning songs about how “there never was a coward where the shamrock grows” and endless paeans to Sikh military prowess.

Sure, scratch the paeans even a little bit and you find condescension as strong as ever. But eight hundred years of the British committing genocide against the Irish and considering them literally subhuman turned into smiles and songs about shamrocks once the Irish started looking like useful cannon fodder for a larger fight. And the Sikhs, dark-skinned people with turbans and beards who pretty much exemplify the European stereotype of “scary foreigner”, were lauded by everyone from the news media all the way up to Winston Churchill.

In other words, outgroups may be the people who look exactly like you, and scary foreigner types can become the in-group on a moment’s notice when it seems convenient.

IV.

There are certain theories of dark matter where it barely interacts with the regular world *at all*, such that we could have a dark matter planet exactly co-incident with Earth and never know. Maybe dark matter people are walking all around us and through us, maybe my house is in the Times Square of a great dark matter city, maybe a few meters away from me a dark matter blogger is writing on his dark matter computer about how weird it would be if there was a light matter person he couldn’t see right next to him.

This is sort of how I feel about conservatives.

I don’t mean the sort of light-matter conservatives who go around complaining about Big Government and occasionally

voting for Romney. I see those guys all the time. What I mean is – well, take creationists. According to Gallup polls, about 46% of Americans are creationists. Not just in the sense of believing God helped guide evolution. I mean they think evolution is a vile atheist lie and God created humans exactly as they exist right now. That's half the country.

And I don't have a *single one of those people* in my social circle. It's not because I'm deliberately avoiding them; I'm pretty live-and-let-live politically, I wouldn't ostracize someone just for some weird beliefs. And yet, even though I probably know about a hundred fifty people, I am pretty confident that not one of them is creationist. Odds of this happening by chance? 1/2^150 = 1/10^45 = approximately the chance of picking a particular atom if you are randomly selecting among all the atoms on Earth.

About forty percent of Americans want to ban gay marriage. I think if I *really* stretch it, maybe ten of my top hundred fifty friends might fall into this group. This is less astronomically unlikely; the odds are a mere one to one hundred quintillion against.

People like to talk about social bubbles, but that doesn't even begin to cover one hundred quintillion. The only metaphor that seems really appropriate is the bizarre dark matter world.

I live in a Republican congressional district in a state with a Republican governor. The conservatives are definitely out there. They drive on the same roads as I do, live in the same neighborhoods. But they might as well be made of dark matter. I never meet them.

To be fair, I spend a lot of my time inside on my computer. I'm browsing sites like Reddit.

Recently, there was a thread on Reddit asking – Redditors Against Gay Marriage, What Is Your Best Supporting Argument? A Reddit user who didn't understand how anybody could be against gay marriage honestly wanted to know how other people who *were* against it justified their position. He figured he might as well ask one of the largest sites on the Internet, with an estimated user base in the tens of millions.

It soon became clear that nobody there was actually against gay marriage.

There were a bunch of posts saying "I of course support gay marriage but here are some reasons some other people might be against it," a bunch of others saying "my argument against gay marriage is the government shouldn't be involved in the marriage business at all", and several more saying "why would you even ask this question, there's no possible good argument and you're wasting your time". About halfway through the thread someone started saying homosexuality was unnatural and I *thought* they were going to be the first one to actually answer the question, but at the end they added "But it's not my place to decide what is or isn't natural, I'm still pro-gay marriage."

In a thread with 10,401 comments, a thread *specifically* asking for people against gay marriage, I was eventually able to find *two* people who came out and opposed it, way near the bottom. Their posts started with "I know I'm going to be downvoted to hell for this…"

But I'm not only on Reddit. I also hang out on LW.

On last year's survey, I found that of American LWers who identify with one of the two major political parties, 80% are Democrat and 20% Republican, which actually sounds pretty balanced compared to some of these other examples.

But it doesn't last. Pretty much all of those "Republicans" are libertarians who consider the GOP the lesser of two evils. When allowed to choose "libertarian" as an alternative, only 4% of visitors continued to identify as conservative. But that's still…some. Right?

When I broke the numbers down further, 3 percentage points of those are neoreactionaries, a bizarre local sect that wants to be ruled by a king. Only *one percent* of LWers were normal everyday God-'n-guns-but-not-George-III conservatives of the type that seem to make up about half of the United States.

It gets worse. My formative years were spent at a university which, if it was similar to other elite universities, had a faculty and a student body that skewed about 90-10 liberal to conservative – and we can bet that, like LW, even those few token conservatives are Mitt Romney types rather than God-n'-guns types. I get my news from vox.com, an Official Liberal Approved Site. Even when I go out to eat, it turns out my favorite restaurant, California Pizza Kitchen, is the most liberal restaurant in the United States.

I inhabit the same geographical area as *scores and scores* of conservatives. But without meaning to, I have created an *outrageously* strong bubble, a 10^45 bubble. Conservatives are all around me, yet I am about as likely to have a serious encounter with one as I am a Tibetan lama.

(Less likely, actually. One time a Tibetan lama came to my college and gave a really nice presentation, but if a conservative tried that, people would protest and it would be canceled.)

V.

One day I realized that entirely by accident I was fulfilling *all* the Jewish stereotypes.

I'm nerdy, over-educated, good with words, good with money, weird sense of humor, don't get outside much, I like deli sandwiches. And I'm a psychiatrist, which is about the most stereotypically Jewish profession short of maybe stand-up comedian or rabbi.

I'm not very religious. And I don't go to synagogue. But *that's* stereotypically Jewish too!

I bring this up because it would be a mistake to think "Well, a Jewish person is by definition someone who is born of a Jewish mother. Or I guess it sort of also means someone who follows the Mosaic Law and goes to synagogue. But I don't care about Scott's mother, and I know he doesn't go to synagogue, so I can't gain any useful information from knowing Scott is Jewish."

The defining factors of Judaism – Torah-reading, synagogue-following, mother-having – are the tip of a giant iceberg. Jews sometimes identify as a "tribe", and even if you don't attend synagogue, you're still a member of that tribe and people can still (in a statistical way) infer things about you by knowing your Jewish identity – like how likely they are to be psychiatrists.

The last section raised a question – if people rarely select their friends and associates and customers explicitly for politics, how do we end up with such intense political segregation?

Well, in the same way "going to synagogue" is merely the iceberg-tip of a Jewish tribe with many distinguishing characteristics, so "voting Republican" or "identifying as conservative" or "believing in creationism" is the iceberg-tip of a conservative tribe with many distinguishing characteristics.

A disproportionate number of my friends are Jewish, because I meet them at psychiatry conferences or something – we self-segregate not based on explicit religion but on implicit tribal characteristics. So in the same way, political tribes self-segregate to an impressive extent – a 1/10^45 extent, I will never tire of hammering in – based on their implicit tribal characteristics.

The people who are actually into this sort of thing sketch out a bunch of speculative tribes and subtribes, but to make it easier, let me stick with two and a half.

The Red Tribe is most classically typified by conservative political beliefs, strong evangelical religious beliefs, creationism, opposing gay marriage, owning guns, eating steak, drinking Coca-Cola, driving SUVs, watching lots of TV, enjoying American football, getting conspicuously upset about terrorists and commies, marrying early, divorcing early, shouting "USA IS NUMBER ONE!!!", and listening to country music.

The Blue Tribe is most classically typified by liberal political beliefs, vague agnosticism, supporting gay rights, thinking guns are barbaric, eating arugula, drinking fancy bottled water, driving Priuses, reading lots of books, being highly educated, mocking American football, feeling vaguely like they should like soccer but never really being able to get into it, getting conspicuously upset about sexists and bigots, marrying later, constantly pointing out how much more civilized European countries are than America, and listening to "everything except country".

(There is a partly-formed attempt to spin off a Grey Tribe typified by libertarian political beliefs, Dawkins-style atheism, vague annoyance that the question of gay rights even comes

up, eating paleo, drinking Soylent, calling in rides on Uber, reading lots of blogs, calling American football "sportsball", getting conspicuously upset about the War on Drugs and the NSA, and listening to filk – but for our current purposes this is a distraction and they can safely be considered part of the Blue Tribe most of the time)

I think these "tribes" will turn out to be even stronger categories than politics. Harvard might skew 80-20 in terms of Democrats vs. Republicans, 90-10 in terms of liberals vs. conservatives, but maybe 99-1 in terms of Blues vs. Reds.

It's the many, many differences between these tribes that explain the strength of the filter bubble – which *have I mentioned* segregates people at a strength of 1/10^45? Even in something as seemingly politically uncharged as going to California Pizza Kitchen or Sushi House for dinner, I'm restricting myself to the set of people who like cute artisanal pizzas or sophsticated foreign foods, which are classically Blue Tribe characteristics.

Are these tribes based on geography? Are they based on race, ethnic origin, religion, IQ, what TV channels you watched as a kid? I don't know.

Some of it is certainly genetic – estimates of the genetic contribution to political association range from 0.4 to 0.6. Heritability of one's attitudes toward gay rights range from 0.3 to 0.5, which hilariously is a little more heritable than homosexuality itself.

(for an interesting attempt to break these down into more rigorous concepts like "traditionalism", "authoritarianism", and "in-group favoritism" and find the genetic loading for each see here. For an attempt to trace the specific genes

involved, which mostly turn out to be NMDA receptors, see here)

But I don't think it's just genetics. There's something else going on too. The word "class" seems like the closest analogue, but only if you use it in the sophisticated Paul Fussell *Guide Through the American Status System* way instead of the boring "another word for how much money you make" way.

For now we can just accept them as a brute fact – as multiple coexisting societies that might as well be made of dark matter for all of the interaction they have with one another – and move on.

VI.

The worst reaction I've ever gotten to a blog post was when I wrote about the death of Osama bin Laden. I've written all sorts of stuff about race and gender and politics and whatever, but that was the worst.

I didn't come out and say I was happy he was dead. But some people interpreted it that way, and there followed a bunch of comments and emails and Facebook messages about how could I possibly be happy about the death of another human being, even if he was a bad person? Everyone, even Osama, is a human being, and we should never rejoice in the death of a fellow man. One commenter came out and said:

> I'm surprised at your reaction. As far as people I casually stalk on the internet (ie, LJ and Facebook), you are the first out of the "intelligent, reasoned and thoughtful" group to be uncomplicatedly happy about this development and not to be, say, disgusted at the reactions of the other 90% or so.

This commenter was right. Of the "intelligent, reasoned, and thoughtful" people I knew, the overwhelming emotion was conspicuous disgust that other people could be happy about his death. I hastily backtracked and said I wasn't happy per se, just surprised and relieved that all of this was finally behind us.

And I genuinely believed that day that I had found some unexpected good in people – that everyone I knew was so humane and compassionate that they were unable to rejoice even in the death of someone who hated them and everything they stood for.

Then a few years later, Margaret Thatcher died. And on my Facebook wall – made of these same "intelligent, reasoned, and thoughtful" people – the most common response was to quote some portion of the song "Ding Dong, The Witch Is Dead". Another popular response was to link the videos of British people spontaneously throwing parties in the street, with comments like "I wish I was there so I could join in". From this exact same group of people, not a single expression of disgust or a "c'mon, guys, we're all human beings here."

I gently pointed this out at the time, and mostly got a bunch of "yeah, so what?", combined with links to an article claiming that "the demand for respectful silence in the wake of a public figure's death is not just misguided but dangerous".

And that was when something clicked for me.

You can talk all you want about Islamophobia, but my friend's "intelligent, reasoned, and thoughtful people" – her name for the Blue Tribe – can't get together enough energy to really hate Osama, let alone Muslims in general. We understand that what he did was bad, but it didn't anger us personally. When he died, we were able to very rationally apply our better nature

and our Far Mode beliefs about how it's never right to be happy about anyone else's death.

On the other hand, that same group absolutely *loathed* Thatcher. Most of us (though not all) can agree, if the question is posed explicitly, that Osama was a worse person than Thatcher. But in terms of actual gut feeling? Osama provokes a snap judgment of "flawed human being", Thatcher a snap judgment of "scum".

I started this essay by pointing out that, despite what geographical and cultural distance would suggest, the Nazis' outgroup was not the vastly different Japanese, but the almost-identical German Jews.

And my hypothesis, stated plainly, is that if you're part of the Blue Tribe, then your outgroup isn't al-Qaeda, or Muslims, or blacks, or gays, or transpeople, or Jews, or atheists – it's the Red Tribe.

VII.

"But racism and sexism and cissexism and anti-Semitism are these giant all-encompassing social factors that verge upon being human universals! Surely you're not arguing that mere *political* differences could ever come close to them!"

One of the ways we *know* that racism is a giant all-encompassing social factor is the Implicit Association Test. Psychologists ask subjects to quickly identify whether words or photos are members of certain gerrymandered categories, like "either a white person's face or a positive emotion" or "either a black person's face and a negative emotion". Then they compare to a different set of gerrymandered categories, like "either a black person's face or a positive emotion" or "either a white person's face or a negative emotion." If subjects have more trouble (as measured in latency time)

connecting white people to negative things than they do white people to positive things, then they probably have subconscious positive associations with white people. You can try it yourself here.

Of course, what the test famously found was that even white people who claimed to have no racist attitudes at all usually had positive associations with white people and negative associations with black people on the test. There are very many claims and counterclaims about the precise meaning of this, but it ended up being a big part of the evidence in favor of the current consensus that all white people are at least a little racist.

Anyway, three months ago, someone finally had the bright idea of doing an Implicit Association Test with political parties, and they found that people's unconscious partisan biases were *half again as strong* as their unconscious racial biases (h/t Bloomberg. For example, if you are a white Democrat, your unconscious bias against blacks (as measured by something called a d-score) is 0.16, but your unconscious bias against Republicans will be 0.23. The Cohen's *d* for racial bias was 0.61, by the book a "moderate" effect size; for party it was 0.95, a "large" effect size.

Okay, fine, but we know race has *real world* consequences. Like, there have been several studies where people sent out a bunch of identical resumes except sometimes with a black person's photo and other times with a white person's photo, and it was noticed that employers were much more likely to invite the fictional white candidates for interviews. So just some stupid Implicit Association Test results can't compare to that, right?

Iyengar and Westwood also decided to do the resume test for parties. They asked subjects to decide which of several candidates should get a scholarship (subjects were told this was a genuine decision for the university the researchers were affiliated with). Some resumes had photos of black people, others of white people. And some students listed their experience in Young Democrats of America, others in Young Republicans of America.

Once again, discrimination on the basis of party was much stronger than discrimination on the basis of race. The size of the race effect for white people was only 56-44 (and in the reverse of the expected direction); the size of the party effect was about 80-20 for Democrats and 69-31 for Republicans.

If you want to see their third experiment, which applied *yet another* classic methodology used to detect racism and *once again* found partyism to be much stronger, you can read the paper.

I & W did an unusually thorough job, but this sort of thing isn't new or ground-breaking. People have been studying "belief congruence theory" – the idea that differences in beliefs are more important than demographic factors in forming in-groups and outgroups – for decades. As early as 1967, Smith et al were doing surveys all over the country and finding that people were more likely to accept friendships across racial lines than across beliefs; in the forty years since then, the observation has been replicated scores of times. Insko, Moe, and Nacoste's 2006 review Belief Congruence And Racial Discrimination concludes that:

> . The literature was judged supportive of a weak version of belief congruence theory which states that in those contexts in which social pressure is nonexistent or

> ineffective, belief is more important than race as a determinant of racial or ethnic discrimination. Evidence for a strong version of belief congruence theory (which states that in those contexts in which social pressure is nonexistent, or ineffective, belief is the only determinant of racial or ethnic discrimination) and was judged much more problematic.

One of the best-known examples of racism is the "Guess Who's Coming To Dinner" scenario where parents are scandalized about their child marrying someone of a different race. Pew has done some good work on this and found that only 23% of conservatives and 1% (!) of liberals admit they would be upset in this situation. But Pew *also* asked how parents would feel about their child marrying someone of a different *political party*. Now 30% of conservatives and 23% of liberals would get upset. Average them out, and you go from 12% upsetness rate for race to 27% upsetness rate for party – more than double. Yeah, people do lie to pollsters, but a picture is starting to come together here.

(Harvard, by the way, is a tossup. There are more black students – 11.5% – than conservative students – 10% – but there are more conservative faculty than black faculty.)

Since people will delight in misinterpreting me here, let me overemphasize what I am *not* saying. I'm not saying people of either party have it "worse" than black people, or that partyism is more of a *problem* than racism, or any of a number of stupid things along those lines which I am sure I will nevertheless be accused of believing. Racism is worse than partyism because the two parties are at least kind of balanced in numbers and in resources, whereas the brunt of an entire country's racism falls on a few underprivileged people. I am saying that the

underlying attitudes that produce partyism are stronger than the underlying attitudes that produce racism, with no necessary implications on their social effects.

But if we want to look at people's psychology and motivations, partyism and the particular variant of tribalism that it represents are going to be fertile ground.

VIII.

Every election cycle like clockwork, conservatives accuse liberals of not being sufficiently pro-America. And every election cycle like clockwork, liberals give extremely unconvincing denials of this.

"It's not that we're, like, *against* America per se. It's just that…well, did you know Europe has much better health care than we do? And much lower crime rates? I mean, come on, how did they get so awesome? And we're just sitting here, can't even get the gay marriage thing sorted out, seriously, what's wrong with a country that can't…sorry, what were we talking about? Oh yeah, America. They're okay. Cesar Chavez was really neat. So were some other people outside the mainstream who became famous precisely by criticizing majority society. That's *sort of* like America being great, in that I think the parts of it that point out how bad the rest of it are often make excellent points. Vote for me!"

(sorry, I make fun of you because I love you)

There was a big brouhaha a couple of years ago when, as it first became apparent Obama had a good shot at the Presidency, Michelle Obama said that "for the first time in my adult life, I am proud of my country."

Republicans pounced on the comment, asking why she hadn't felt proud before, and she backtracked saying of course she

was proud all the time and she loves America with the burning fury of a million suns and she was just saying that the Obama campaign was *particularly* inspiring.

As unconvincing denials go, this one was pretty far up there. But no one really held it against her. Probably most Obama voters felt vaguely the same way. *I* was an Obama voter, and I have proud memories of spending my Fourth of Julys as a kid debunking people's heartfelt emotions of patriotism. Aaron Sorkin:

> [What makes America the greatest country in the world?] It's not the greatest country in the world! We're seventh in literacy, 27th in math, 22nd in science, 49th in life expectancy, 178th in infant mortality, third in median household income, No. 4 in labor force, and No. 4 in exports. So when you ask what makes us the greatest country in the world, I don't know what the f*** you're talking about.

(Another good retort is "We're number one? Sure – number one in incarceration rates, drone strikes, and making new parents go back to work!")

All of this is true, of course. But it's weird that it's such a classic interest of members of the Blue Tribe, and members of the Red Tribe never seem to bring it up.

("We're number one? Sure – number one in levels of sexual degeneracy! Well, I guess probably number two, after the Netherlands, but they're really small and shouldn't count.")

My hunch – both the Red Tribe and the Blue Tribe, for whatever reason, identify "America" with the Red Tribe. Ask people for typically "American" things, and you end up with a very Red list of characteristics – guns, religion, barbecues,

American football, NASCAR, cowboys, SUVs, unrestrained capitalism.

That means the Red Tribe feels intensely patriotic about "their" country, and the Blue Tribe feels like they're living in fortified enclaves deep in hostile territory.

Here is a popular piece published on a major media site called America: A Big, Fat, Stupid Nation. Another: America: A Bunch Of Spoiled, Whiny Brats. Americans are ignorant, scientifically illiterate religious fanatics whose "patriotism" is actually just narcissism. You Will Be Shocked At How Ignorant Americans Are, and we should Blame The Childish, Ignorant American People.

Needless to say, every single one of these articles was written by an American and read almost entirely by Americans. Those Americans very likely enjoyed the articles very much and did not feel the least bit insulted.

And look at the sources. HuffPo, Salon, Slate. Might those have anything in common?

On both sides, "American" can be either a normal demonym, or a code word for a member of the Red Tribe.

IX.

The other day, I logged into OKCupid and found someone who looked cool. I was reading over her profile and found the following sentence:

> Don't message me if you're a sexist white guy

And my first thought was "Wait, so a sexist black person would be okay? Why?"

(The girl in question was white as snow)

Around the time the Ferguson riots were first starting, there were a host of articles with titles like Why White People Don't Seem To Understand Ferguson, Why It's So Hard For Whites To Understand Ferguson, and White Folks Listen Up And Let Me Tell You What Ferguson Is All About, this last of which says:

> Social media is full of people on both sides making presumptions, and believing what they want to believe. But it's the white folks that don't understand what this is all about. Let me put it as simply as I can for you […]
>
> No matter how wrong you think Trayvon Martin or Michael Brown were, I think we can all agree they didn't deserve to die over it. I want you white folks to understand that this is where the anger is coming from. You focused on the looting…."

And on a hunch I checked the author photos, and every single one of these articles was written by a white person.

White People Are Ruining America? White. White People Are Still A Disgrace? White. White Guys: We Suck And We're Sorry? White. Bye Bye, Whiny White Dudes? White. Dear Entitled Straight White Dudes, I'm Evicting You From My Life? White. White Dudes Need To Stop Whitesplaining? White. Reasons Why Americans Suck #1: White People? White.

We've all seen articles and comments and articles like this. Some unsavory people try to use them to prove that white people are the *real* victims or the media is biased against white people or something. Other people who are very nice and optimistic use them to show that some white people have

developed some self-awareness and are willing to engage in self-criticism.

But I think the situation with "white" is much the same as the situation with "American" – it can either mean what it says, or be a code word for the Red Tribe.

(except on the blog Stuff White People Like, where it obviously serves as a code word for the *Blue* tribe. I don't know, guys. I didn't do it.)

I realize that's making a strong claim, but it would hardly be without precedent. When people say things like "gamers are misogynist", do they mean the 52% of gamers who are women? Do they mean every one of the 59% of Americans from every walk of life who are known to play video or computer games occasionally? No. "Gamer" is a coded reference to the Gray Tribe, the half-branched-off collection of libertarianish tech-savvy nerds, and everyone knows it. As well expect that when people talk about "fedoras", they mean Indiana Jones. Or when they talk about "urban youth", they mean freshmen at NYU. Everyone knows exactly who we mean when we say "urban youth", and them being young people who live in a city has only the most tenuous of relations to the actual concept.

And I'm saying words like "American" and "white" work the same way. Bill Clinton was the "first black President", but if Herman Cain had won in 2012 he'd have been the 43rd white president. And when an angry white person talks at great length about how much he hates "white dudes", *he is not being humble and self-critical*.

X.

Imagine hearing that a liberal talk show host and comedian was so enraged by the actions of ISIS that he'd recorded and

posted a video in which he shouts at them for ten minutes, cursing the "fanatical terrorists" and calling them "utter savages" with "savage values".

If *I* heard that, I'd be kind of surprised. It doesn't fit my model of what liberal talk show hosts do.

But the story I'm *actually* referring to is liberal talk show host / comedian Russell Brand making that same rant against Fox News for *supporting war against* the Islamic State, adding at the end that "Fox is worse than ISIS".

That fits my model perfectly. You wouldn't celebrate Osama's death, only Thatcher's. And you wouldn't call ISIS savages, only Fox News. Fox is the outgroup, ISIS is just some random people off in a desert. You hate the outgroup, you don't hate random desert people.

I would go further. Not only does Brand not feel much like hating ISIS, he has a strong incentive not to. That incentive is: the Red Tribe is known to hate ISIS loudly and conspicuously. Hating ISIS would signal Red Tribe membership, would be the equivalent of going into Crips territory with a big Bloods gang sign tattooed on your shoulder.

But this might be unfair. What would Russell Brand answer, if we asked him to justify his decision to be much angrier at Fox than ISIS?

He might say something like "Obviously Fox News is not literally worse than ISIS. But here I am, talking to my audience, who are mostly white British people and Americans. These people already know that ISIS is bad; they don't need to be told that any further. In fact, at this point being angry about how bad ISIS is, is less likely to genuinely change someone's mind about ISIS, and more likely to promote Islamophobia. The sort of people in my audience are at zero risk of becoming

ISIS supporters, but at a very real risk of Islamophobia. So ranting against ISIS would be counterproductive and dangerous.

On the other hand, my audience of white British people and Americans is very likely to contain many Fox News viewers and supporters. And Fox, while not quite as evil as ISIS, is still pretty bad. So here's somewhere I have a genuine chance to reach people at risk and change minds. Therefore, I think my decision to rant against Fox News, and maybe hyperbolically say they were 'worse than ISIS' is justified under the circumstances."

I have a lot of sympathy to hypothetical-Brand, especially to the part about Islamophobia. It *does* seem really possible to denounce ISIS' atrocities to a population that already hates them in order to weak-man a couple of already-marginalized Muslims. We need to fight terrorism and atrocities – therefore it's okay to shout at a poor girl ten thousand miles from home for wearing a headscarf in public. Christians are being executed for their faith in Sudan, therefore let's picket the people trying to build a mosque next door.

But my sympathy with Brand ends when he acts like his audience is likely to be fans of Fox News.

In a world where a negligible number of Redditors oppose gay marriage and 1% of Less Wrongers identify conservative and I know 0/150 creationists, how many of the people who visit the YouTube channel of a well-known liberal activist with a Che-inspired banner, a channel whose episode names are things like "War: What Is It Good For?" and "Sarah Silverman Talks Feminism" – how many of them do you think are big Fox News fans?

In a way, Russell Brand would have been *braver* taking a stand against ISIS than against Fox. If he attacked ISIS, his viewers would just be a little confused and uncomfortable. Whereas every moment he's attacking Fox his viewers are like "HA HA! YEAH! GET 'EM! SHOW THOSE IGNORANT BIGOTS IN THE outgroup WHO'S BOSS!"

Brand acts as if there are just these countries called "Britain" and "America" who are receiving his material. Wrong. There are two parallel universes, and he's only broadcasting to one of them.

The result is exactly what we predicted would happen in the case of Islam. Bombard people with images of a far-off land they already hate and tell them to hate it more, and the result is ramping up the intolerance on the couple of dazed and marginalized representatives of that culture who have ended up stuck on your half of the divide. Sure enough, if industry or culture or community gets Blue enough, Red Tribe members start getting harassed, fired from their jobs (Brendan Eich being the obvious example) or otherwise shown the door.

Think of Brendan Eich as a member of a tiny religious minority surrounded by people who hate that minority. Suddenly firing him doesn't seem very noble.

If you mix together Podunk, Texas and Mosul, Iraq, you can prove that Muslims are scary and very powerful people who are executing Christians all the time and have a great excuse for kicking the one remaining Muslim family, random people who never hurt anyone, out of town.

And if you mix together the open-source tech industry and the parallel universe where you can't wear a FreeBSD t-shirt without risking someone trying to exorcise you, you can prove that Christians are scary and very powerful people who are

persecuting everyone else all the time, and you have a great excuse for kicking one of the few people willing to affiliate with the Red Tribe, a guy who never hurt anyone, out of town.

When a friend of mine heard Eich got fired, she didn't see anything wrong with it. "I can tolerate anything except intolerance," she said.

"Intolerance" is starting to look like another one of those words like "white" and "American".

"I can tolerate anything except the outgroup." Doesn't sound quite so noble now, does it?

XI.

We started by asking: millions of people are conspicuously praising every outgroup they can think of, while conspicuously condemning their own in-group. This seems contrary to what we know about social psychology. What's up?

We noted that outgroups are rarely literally "the group most different from you", and in fact far more likely to be groups very similar to you sharing *almost* all your characteristics and living in the same area.

We then noted that although liberals and conservatives live in the same area, they might as well be two totally different countries or universe as far as level of interaction were concerned.

Contra the usual idea of them being marked only by voting behavior, we described them as very different tribes with totally different cultures. You can speak of "American culture" only in the same way you can speak of "Asian culture" – that is, with a lot of interior boundaries being pushed under the rug.

The outgroup of the Red Tribe is occasionally blacks and gays and Muslims, more often the Blue Tribe.

The Blue Tribe has performed some kind of very impressive act of alchemy, and transmuted *all* of its outgroup hatred to the Red Tribe.

This is not surprising. Ethnic differences have proven quite tractable in the face of shared strategic aims. Even the Nazis, not known for their ethnic tolerance, were able to get all buddy-buddy with the Japanese when they had a common cause.

Research suggests Blue Tribe / Red Tribe prejudice to be much stronger than better-known types of prejudice like racism. Once the Blue Tribe was able to enlist the blacks and gays and Muslims in their ranks, they became allies of convenience who deserve to be rehabilitated with mildly condescending paeans to their virtue. "There never was a coward where the shamrock grows."

Spending your entire life insulting the other tribe and talking about how terrible they are makes you look, well, tribalistic. It is definitely not high class. So when members of the Blue Tribe decide to dedicate their entire life to yelling about how terrible the Red Tribe is, they make sure that instead of saying "the Red Tribe", they say "America", or "white people", or "straight white men". That way it's *humble self-criticism*. They are *so* interested in justice that they are willing to critique *their own beloved side*, much as it pains them to do so. We know they are not exaggerating, because one might exaggerate the flaws of an enemy, but that anyone would exaggerate their *own* flaws fails the criterion of embarrassment.

The Blue Tribe always has an excuse at hand to persecute and crush any Red Tribers unfortunate enough to fall into its light-matter-universe by defining them as all-powerful domineering

oppressors. They appeal to the fact that this is definitely the way it works in the Red Tribe's dark-matter-universe, and that's in the same country so it has to be the same community for all intents and purposes. As a result, every Blue Tribe institution is permanently licensed to take whatever emergency measures are necessary against the Red Tribe, however disturbing they might otherwise seem.

And so how virtuous, how noble the Blue Tribe! Perfectly tolerant of all of the different groups that just so happen to be allied with them, never intolerant unless it happen to be against intolerance itself. Never stooping to engage in petty tribal conflict like that awful Red Tribe, but always nobly criticizing their own culture and striving to make it better!

Sorry. But I hope this is at least a *little* convincing. The weird dynamic of outgroup-philia and ingroup-phobia isn't anything of the sort. It's just good old-fashioned in-group-favoritism and outgroup bashing, a little more sophisticated and a little more sneaky.

XII.

This essay is bad and I should feel bad.

I should feel bad because I made *exactly* the mistake I am trying to warn everyone else about, and it wasn't until I was almost done that I noticed.

How virtuous, how noble I must be! Never stooping to engage in petty tribal conflict like that silly Red Tribe, but always nobly criticizing my own tribe and striving to make it better.

Yeah. Once I've written a ten thousand word essay savagely attacking the Blue Tribe, either I'm a very special person or they're my outgroup. And I'm not *that* special.

Just as you can pull a fast one and look humbly self-critical if you make your audience assume there's just one American culture, so maybe you can trick people by assuming there's only one Blue Tribe.

I'm pretty sure I'm not Red, but I did talk about the Grey Tribe above, and I show all the risk factors for being one of them. That means that, although my critique of the Blue Tribe may be right or wrong, in terms of *motivation* it comes from the same place as a Red Tribe member talking about how much they hate al-Qaeda or a Blue Tribe member talking about how much they hate ignorant bigots. And when I boast of being able to tolerate Christians and Southerners whom the Blue Tribe is mean to, I'm not being tolerant at all, just noticing people so far away from me they wouldn't make a good outgroup anyway.

My arguments might be *correct* feces, but they're still feces.

I had *fun* writing this article. People do not have fun writing articles savagely criticizing their in-group. People can criticize their in-group, it's not *humanly impossible*, but it takes nerves of steel, it makes your blood boil, you should sweat blood. It shouldn't be *fun*.

You can bet some white guy on Gawker who week after week churns out "Why White People Are So Terrible" and "Here's What Dumb White People Don't Understand" is having fun and not sweating any blood at all. He's not criticizing his in-group, he's never even *considered* criticizing his in-group. I can't blame him. Criticizing the in-group is a really difficult project I've barely begun to build the mental skills necessary to even consider.

I can think of criticisms of my own tribe. Important criticisms, true ones. But the thought of writing them makes my blood

boil.

I imagine might I feel like some liberal US Muslim leader, when he goes on the O'Reilly Show, and O'Reilly ambushes him and demands to know why he and other American Muslims haven't condemned beheadings by ISIS more, demands that he criticize them right there on live TV. And you can see the wheels in the Muslim leader's head turning, thinking something like "Okay, obviously beheadings are terrible and I hate them as much as anyone. But you don't care even *the slightest bit* about the victims of beheadings. You're just looking for a way to score points against me so you can embarass all Muslims. And I would rather personally behead every single person in the world than give a smug bigot like you a single microgram more stupid self-satisfaction than you've already got."

That is how I feel when asked to criticize my own tribe, even for correct reasons. If you think you're criticizing your own tribe, and your blood is not at that temperature, consider the possibility that you aren't.

But if I want Self-Criticism Virtue Points, criticizing the Grey Tribe is the only honest way to get them. And if I want Tolerance Points, my own personal cross to bear right now is tolerating the Blue Tribe. I need to remind myself that when they are bad people, they are merely Osama-level bad people instead of Thatcher-level bad people. And when they are good people, they are powerful and necessary crusaders against the evils of the world.

The worst thing that could happen to this post is to have it be used as convenient feces to fling at the Blue Tribe whenever feces are necessary. Which, given what has happened to my

last couple of posts along these lines and the obvious biases of my own subconscious, I already expect it will be.

But the best thing that could happen to this post is that it makes a lot of people, especially myself, figure out how to be more tolerant. Not in the "of course I'm tolerant, why shouldn't I be?" sense of the Emperor in Part I. But in the sense of "being tolerant makes me see red, makes me sweat blood, but darn it *I am going to be tolerant anyway*."

Five Case Studies on Politicization

[Trigger warning: Some discussion of rape in Part III. This will make much more sense if you've previously read I Can Tolerate Anything Except The Outgroup]

I.

One day I woke up and they had politicized Ebola.

I don't just mean the usual crop of articles like Republicans Are Responsible For The Ebola Crisis and Democrats Try To Deflect Blame For Ebola Outbreak and Incredibly Awful Democrats Try To Blame Ebola On GOP and NPR Reporter Exposes Right Wing Ebola Hype and Republicans Flip-Flop On Ebola Czars. That level of politicization was pretty much what I *expected*.

(I can't say I *totally* expected to see an article called Fat Lesbians Got All The Ebola Dollars, But Blame The GOP, but in retrospect nothing I know about modern society suggested I wouldn't)

I'm talking about something weirder. Over the past few days, my friends on Facebook have been making impassioned posts about how it's *obvious* there should/shouldn't be a quarantine, but deluded people on the other side are muddying the issue. The issue has risen to an alarmingly high level of 0.05 #Gamergates, which is my current unit of how much people on social media are concerned about a topic. What's more, everyone supporting the quarantine has been on the right, and everyone opposing on the left. *Weird* that so many people suddenly develop strong feelings about a complicated epidemiological issue, which can be exactly predicted by their feelings about *everything else*.

On the Right, there is condemnation of the CDC's opposition to quarantines as globalist gibberish, fourteen questions that

will never be asked about Ebola centering on why there aren't more quarantine measures in place, and arguments on right-leaning biology blogs for why the people opposing quarantines are dishonest or incompetent. Top Republicans call for travel bans and a presenter on Fox, proportionate as always, demands quarantine centers in every US city.

On the Left (and token libertarian) sides, the New Yorker has been publishing articles on how involuntary quarantines violate civil liberties and "embody class and racial biases", Reason makes fun of "dumb Republican calls for a travel ban", Vox has a clickbaity article on how "This One Paragraph Perfectly Sums Up America's Overreaction To Ebola", and MSNBC notes that to talk about travel bans is "borderline racism".

How did this happen? How did both major political tribes decide, within a month of the virus becoming widely known in the States, not only exactly what their position should be but what insults they should call the other tribe for not agreeing with their position? There are a lot of complicated and well-funded programs in West Africa to disseminate information about the symptoms of Ebola in West Africa, and all I can think of right now is that if the Africans could disseminate useful medical information half as quickly as Americans seem to have disseminated tribal-affiliation-related information, the epidemic would be over tomorrow.

Is it just random? A couple of Republicans were coincidentally the first people to support a quarantine, so other Republicans felt they had to stand by them, and then Democrats felt they had to oppose it, and then that spread to wider and wider circles? And if by chance a Democrats had proposed quarantine before a Republican, the situation would have reversed itself? Could be.

Much more interesting is the theory that the fear of disease is the root of all conservativism. I am not making this up. There has been a lot of really good evolutionary psychology done on the extent to which pathogen stress influences political opinions. Some of this is done on the societal level, and finds that societies with higher germ loads are more authoritarian and conservative. This research can be followed arbitrarily far – like, isn't it *interesting* that the most liberal societies in the world are the Scandinavian countries in the very far north where disease burden is low, and the most traditionalist-authoritarian ones usually in Africa or somewhere where disease burden is high? One even sees a similar effect within countries, with northern US states being very liberal and southern states being very conservative. Other studies have instead focused on differences between individuals within society – we know that religious conservatives are people with stronger disgust reactions and priming disgust reactions can increase self-reported conservative political beliefs – with most people agreeing disgust reactions are a measure of the "behavioral immune system" triggered by fear of germ contamination.

(free tip for liberal political activists – offering to tidy up voting booths before the election is probably a thousand times more effective than anything you're doing right now. I will leave the free tip for conservative political activists to your imagination)

If being a conservative means you're pre-selected for worry about disease, obviously the conservatives are going to be the ones most worried about Ebola. And in fact, along with the quarantine debate, there's a little sub-debate about whether Ebola is worth panicking about. Vox declares Americans to be "overreacting" and keeps telling them to calm down, whereas

its similarly-named evil twin Vox Day has been spending the last week or so spreading panic and suggesting readers "wash your hands, stock up a bit, and avoid any unnecessary travel".

So that's the second theory.

The third theory is that everything in politics is mutually reinforcing.

Suppose the Red Tribe has a Grand Narrative. The Narrative is something like *"We Americans are right-thinking folks with a perfectly nice culture. But there are also scary foreigners who hate our freedom and wish us ill. Unfortunately, there are also traitors in our ranks – in the form of the Blue Tribe – who in order to signal sophistication support foreigners over Americans and want to undermine our culture. They do this by supporting immigration, accusing anyone who is too pro-American and insufficiently pro-foreigner of "racism", and demanding everyone conform to "multiculturalism" and "diversity", as well as lionizing any group within America that tries to subvert the values of the dominant culture. Our goal is to minimize the subversive power of the Blue Tribe at home, then maintain isolation from foreigners abroad, enforced by a strong military if they refuse to stay isolated."*

And the Blue Tribe also has a Grand Narrative. The Narrative is something like *"The world is made up of a bunch of different groups and cultures. The wealthier and more privileged groups, played by the Red Tribe, have a history of trying to oppress and harass all the other groups. This oppression is based on ignorance, bigotry, xenophobia, denial of science, and a false facade of patriotism. Our goal is to call out the Red Tribe on its many flaws, and support other groups like foreigners and minorities in their quest for justice and*

equality, probably in a way that involves lots of NGOs and activists."

The proposition "a quarantine is the best way to deal with Ebola" seems to fit much better into the Red narrative than the Blue Narrative. It's about foreigners being scary and dangerous, and a strong coordinated response being necessary to protect right-thinking Americans from them. When people like NBC and the New Yorker accuse quarantine opponents of being "racist", that just makes the pieces fit in all the better.

The proposition "a quarantine is a bad way to deal with Ebola" seems to fit much better into the Blue narrative than the Red. It's about extremely poor black foreigners dying, and white Americans rushing to throw them overboard to protect themselves out of ignorance of the science (which says Ebola can't spread much in the First World), bigotry, xenophobia, and fear. The *real* solution is a coordinated response by lots of government agencies working in tandem with NGOs and local activists.

It would be really hard to switch these two positions around. If the Republicans were to oppose a quarantine, it might raise the *general question* of whether closing the borders and being scared of foreign threats is always a good idea, and whether maybe sometimes accusations of racism are making a good point. Far "better" to maintain a consistent position where all your beliefs reinforce all of your other beliefs.

There's a question of causal structure here. Do Republicans believe certain other things for their own sake, and then adapt their beliefs about Ebola to help buttress their other beliefs? Or do the same factors that made them adopt their narrative in the first place lead them to adopt a similar narrative around Ebola?

My guess it it's a little of both. And then once there's a critical mass of anti-quarantiners within a party, in-group cohesion and identification effects cascade towards it being a badge of party membership and everybody having to believe it. And if the Democrats are on the other side, saying things you disagree with about every *other* issue, and also saying that you have to oppose quarantine or else you're a bad person, then that also incentivizes you to support a quarantine, *just to piss them off*.

II.

Sometimes politicization isn't about what side you take, it's about what issues you emphasize.

In the last post, I wrote:

> Imagine hearing that a liberal talk show host and comedian was so enraged by the actions of ISIS that he'd recorded and posted a video in which he shouts at them for ten minutes, cursing the "fanatical terrorists" and calling them "utter savages" with "savage values".
>
> If I heard that, I'd be kind of surprised. It doesn't fit my model of what liberal talk show hosts do.
>
> But the story I'm actually referring to is liberal talk show host / comedian Russell Brand making that same rant against Fox News for supporting war against the Islamic State, adding at the end that "Fox is worse than ISIS".
>
> That fits my model perfectly. You wouldn't celebrate Osama's death, only Thatcher's. And you wouldn't call ISIS savages, only Fox News. Fox is the outgroup, ISIS is just some random people off in a desert. You hate the outgroup, you don't hate random desert people.

> I would go further. Not only does Brand not feel much like hating ISIS, he has a strong incentive not to. That incentive is: the Red Tribe is known to hate ISIS loudly and conspicuously. Hating ISIS would signal Red Tribe membership, would be the equivalent of going into Crips territory with a big Bloods gang sign tattooed on your shoulder.

Now I think I missed an important part of the picture. The existence of ISIS plays right into Red Tribe narratives. They are *totally* scary foreigners who hate our freedom and want to hurt us and probably require a strong military response, so their existence sounds like a point in favor of the Red Tribe. Thus, the Red Tribe wants to talk about them as much as possible and condemn them in the strongest terms they can.

There's not really any way to spin this issue in favor of the Blue Tribe narrative. The Blue Tribe just has to grudgingly admit that maybe this is one of the few cases where their narrative breaks down. So their incentive is to try to minimize ISIS, to admit it exists and is bad and try to distract the conversation to other issues that support their chosen narrative more. That's why you'll never see the Blue Tribe gleefully cheering someone on as they call ISIS "savages". It wouldn't fit the script.

But did you hear about that time when a Muslim-American lambasting Islamophobia totally pwned all of those ignorant FOX anchors? Le-GEN-dary!

III.

At worst this choice to emphasize different issues descends into an unhappy combination of tragedy and farce.

The Rotherham scandal was an incident in an English town where criminal gangs had been grooming and blackmailing thousands of young girls, then using them as sex slaves. This had been going on for at least ten years with minimal intervention by the police. An investigation was duly launched, which discovered that the police had been keeping quiet about the problem because the gangs were mostly Pakistani and the victims mostly white, and the police didn't want to seem racist by cracking down too heavily. Researchers and officials who demanded that the abuse should be publicized or fought more vigorously were ordered to attend "diversity training" to learn why their demands were offensive. The police department couldn't keep it under wraps forever, and eventually it broke and was a huge scandal.

The Left then proceeded to totally ignore it, and the Right proceeded to *never shut up* about it for like an entire month, and every article about it had to include the "diversity training" aspect, so that if you type "rotherham d…" into Google, your two first options are "Rotherham Daily Mail" and "Rotherham diversity training".

I don't find this surprising at all. The Rotherham incident ties in *perfectly* to the Red Tribe narrative – scary foreigners trying to hurt us, politically correct traitors trying to prevent us from noticing. It doesn't do *anything* for the Blue Tribe narrative, and indeed actively contradicts it at some points. So the Red Tribe wants to trumpet it to the world, and the Blue Tribe wants to stay quiet and distract.

HBD Chick usually writes very well-thought-out articles on race and genetics listing all the excellent reasons you should not marry your cousins. Hers is not a political blog, and I have never seen her get upset about any political issue before, but since most of her posts are about race and genetics she gets a

lot of love from the Right and a lot of flak from the Left. She recently broke her silence on politics to write three long and very angry blog posts on the Rotherham issue, of which I will excerpt one:

> if you've EVER called somebody a racist just because they said something politically incorrect, then you'd better bloody well read this report, because THIS IS ON YOU! this is YOUR doing! this is where your scare tactics have gotten us: over 1400 vulnerable kids systematically abused because YOU feel uncomfortable when anybody brings up some "hate facts."
>
> this is YOUR fault, politically correct people — and i don't care if you're on the left or the right. YOU enabled this abuse thanks to the climate of fear you've created. thousands of abused girls — some of them maybe dead — on YOUR head.

I have no doubt that her outrage is genuine. But I do have to wonder why she is outraged about this and not all of the other outrageous things in the world. And I do have to wonder whether the perfect fit between her own problems – trying to blog about race and genetics but getting flak from politically correct people – and the problems that made Rotherham so disastrous – which include police getting flak from politically correct people – are part of her sudden conversion to political activism.

[edit: she objects to this characterization]

But I will also give her this – accidentally stumbling into being upset by the rape of thousands of children is, as far as accidental stumbles go, not a bad one. What's everyone *else's* excuse?

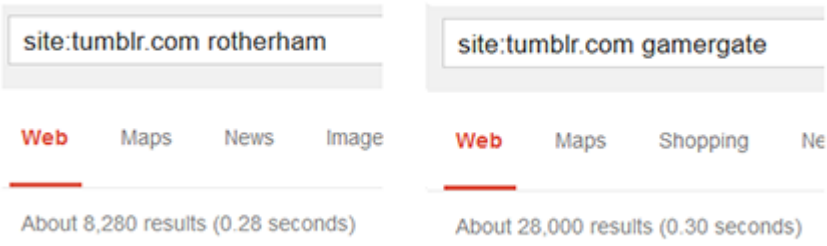
John Durant did an interesting analysis of media coverage of the Rotherham scandal versus the "someone posted nude pictures of Jennifer Lawrence" scandal.

He found left-leaning news website Slate had one story on the Rotherham child exploitation scandal, but four stories on nude Jennifer Lawrence.

He also found that feminist website Jezebel had only one story on the Rotherham child exploitation scandal, but six stories on nude Jennifer Lawrence.

Feministing gave Rotherham a one-sentence mention in a links roundup (just underneath "five hundred years of female portrait painting in three minutes"), but Jennifer Lawrence got two full stories.

The article didn't talk about social media, and I couldn't search it directly for Jennifer Lawrence stories because it was too hard to sort out discussion of the scandal from discussion of her as an actress. But using my current unit of social media saturation, Rotherham clocks in at 0.24 #Gamergates



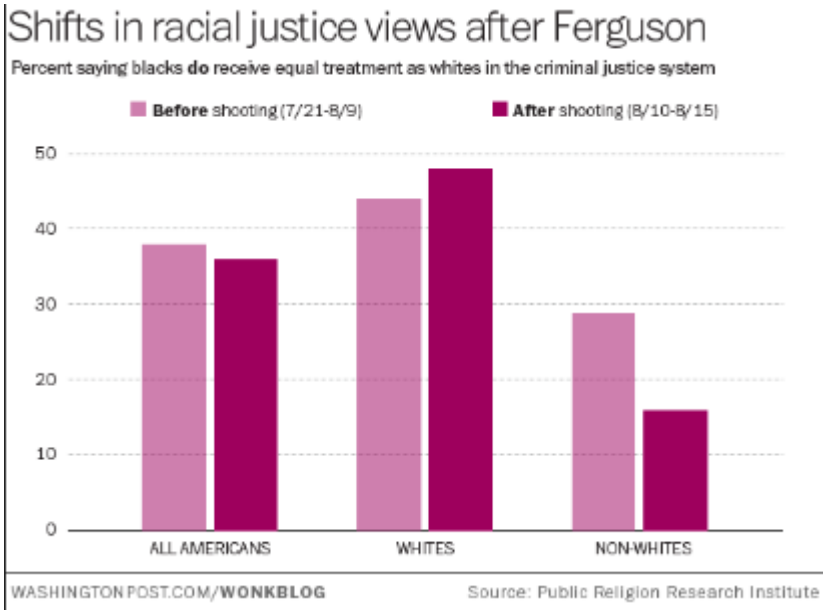
You thought I was joking. I never joke.

This doesn't surprise me much. Yes, you would think that the systematic rape of thousands of women with police taking no action might be a feminist issue. Or that it might outrage some people on Tumblr, a site which has many flaws but which has never been accused of being slow to outrage. But the goal here isn't to push some kind of Platonic ideal of what's important,

it's to support a certain narrative that ties into the Blue Tribe narrative. Rotherham does the opposite of that. The Jennifer Lawrence nudes, which center around how hackers (read: creepy internet nerds) shared nude pictures of a beloved celebrity on Reddit (read: creepy internet nerds) and 4Chan (read: creepy internet nerds) – and #Gamergate which does the same – are *exactly* the narrative they want to push, so they become the Stories Of The Century.

IV.

Here's something I *did* find on Tumblr which I think is really interesting.



You can see that after the Ferguson shooting, the average American became a little less likely to believe that blacks were treated equally in the criminal justice system. This makes sense, since the Ferguson shooting was a much-publicized example of the criminal justice system treating a black person unfairly.

But when you break the results down by race, a different picture emerges. White people were actually a little *more* likely to believe the justice system was fair after the shooting.

Why? I mean, if there was no change, you could chalk it up to white people believing the police's story that the officer involved felt threatened and made a split-second bad decision that had nothing to do with race. That could explain no change just fine. But being *more* convinced that justice is color-blind? What could explain *that*?

My guess – before Ferguson, at least a few people interpreted this as an honest question about race and justice. After Ferguson, everyone mutually agreed it was about politics.

Ferguson and Rotherham were both similar in that they were cases of police misconduct involving race. You would think that there might be some police misconduct community who are interested in stories of police misconduct, or some race community interested in stories about race, and these people would discuss both of these two big international news items.

The Venn diagram of sources I saw covering these two stories forms two circles with no overlap. All those conservative news sites that couldn't shut up about Rotherham? Nothing on Ferguson – unless it was to snipe at the Left for "exploiting" it to make a political point. Otherwise, they did their best to stay quiet about it. Hey! Look over there! ISIS is probably beheading someone really interesting!

The same way Rotherham obviously supports the Red Tribe's narrative, Ferguson obviously supports the Blue Tribe's narrative. A white person, in the police force, shooting an innocent (ish) black person, and then a racist system refusing to listen to righteous protests by brave activists.

The "see, the Left is right about everything" angle of most of the coverage made HBD Chick's attack on political correctness look subtle. The parts about race, systemic inequality, and the police were of debatable proportionality,

but what I really liked was the Ferguson coverage started branching off into every issue any member of the Blue Tribe has ever cared about:

Gun control? Check.

The war on terror? Check.

American exceptionalism? Check.

Feminism? Check.

Abortion? Check

Gay rights? Check.

Palestinian independence? Check.

Global warming? Check. Wait, really? Yes, really.

Anyone who thought that the question in that poll was just a simple honest question about criminal justice was very quickly disabused of that notion. It was a giant Referendum On Everything, a "do you think the Blue Tribe is right on every issue and the Red Tribe is terrible and stupid, or vice versa?" And it turns out many people who when asked about criminal justice will just give the obvious answer, have much stronger and less predictable feelings about Giant Referenda On Everything.

In my last post, I wrote about how people feel when their in-group is threatened, even when it's threatened with an apparently innocuous point they totally agree with:

> I imagine [it] might feel like some liberal US Muslim leader, when he goes on the O'Reilly Show, and O'Reilly ambushes him and demands to know why he and other American Muslims haven't condemned beheadings by ISIS more, demands that he criticize them right there on

> live TV. And you can see the wheels in the Muslim leader's head turning, thinking something like "Okay, obviously beheadings are terrible and I hate them as much as anyone. But you don't care even the slightest bit about the victims of beheadings. You're just looking for a way to score points against me so you can embarass all Muslims. And I would rather personally behead every single person in the world than give a smug bigot like you a single microgram more stupid self-satisfaction than you've already got."

I think most people, when they think about it, probably believe that the US criminal justice system is biased. But when you feel under attack by people whom you suspect have dishonest intentions of twisting your words so they can use them to dehumanize your in-group, eventually you think "I would rather personally launch unjust prosecutions against every single minority in the world than give a smug out-group member like you a single microgram more stupid self-satisfaction than you've already got."

V.

Wait, so you mean turning all the most important topics in our society into wedge issues that we use to insult and abuse people we don't like, to the point where even mentioning it triggers them and makes them super defensive, might have been a *bad* idea??!

There's been some really neat research into people who don't believe in global warming. The original suspicion, at least from certain quarters, were that they were just dumb. Then someone checked and found that warming disbelievers actually had (very slightly) higher levels of scientific literacy than warming believers.

So people had to do actual studies, and to what should have been no one's surprise, the most important factor was partisan affiliation. For example, according to Pew 64% of Democrats believe the Earth is getting warmer due to human activity, compared to 9% of Tea Party Republicans.

So assuming you want to convince Republicans to start believing in global warming before we're all frying eggs on the sidewalk, how should you go about it? This is the excellent question asked by a study recently profiled in an NYMag article.

The study found that you could be a little more convincing to conservatives by acting on the purity/disgust axis of moral foundations theory – the one that probably gets people so worried about Ebola. A warmer climate is *unnatural*, in the same way that, oh, let's say, homosexuality is unnatural. Carbon dioxide *contaminating* our previously pure atmosphere, in the same way premarital sex or drug use contaminates your previously pure body. It sort of worked.

Another thing that sort of worked was *tying things into the Red Tribe narrative*, which they did through the two sentences "Being pro-environmental allows us to protect and preserve the American way of life. It is patriotic to conserve the country's natural resources." I can't imagine anyone falling for this, but I guess some people did.

This is cute, but it's too little too late. Global warming has already gotten inextricably tied up in the Blue Tribe narrative: *Global warming proves that unrestrained capitalism is destroying the planet. Global warming disproportionately affects poor countries and minorities. Global warming could have been prevented with multilateral action, but we were too dumb to participate because of stupid American cowboy*

diplomacy. Global warming is an important cause that activists and NGOs should be lauded for highlighting. Global warming shows that Republicans are science denialists and probably all creationists. Two lousy sentences on "patriotism" aren't going to break through that.

If *I* were in charge of convincing the Red Tribe to line up behind fighting global warming, here's what I'd say:

> In the 1950s, brave American scientists shunned by the climate establishment of the day discovered that the Earth was warming as a result of greenhouse gas emissions, leading to potentially devastating natural disasters that could destroy American agriculture and flood American cities. As a result, the country mobilized against the threat. Strong government action by the Bush administration outlawed the worst of these gases, and brilliant entrepreneurs were able to discover and manufacture new cleaner energy sources. As a result of these brave decisions, our emissions stabilized and are currently declining.
>
> Unfortunately, even as we do our part, the authoritarian governments of Russia and China continue to industralize and militarize rapidly as part of their bid to challenge American supremacy. As a result, Communist China is now by far the world's largest greenhouse gas producer, with the Russians close behind. Many analysts believe Putin secretly welcomes global warming as a way to gain access to frozen Siberian resources and weaken the more temperate United States at the same time. These countries blow off huge disgusting globs of toxic gas, which effortlessly cross American borders and disrupt the climate of the United States. Although we have asked

them to stop several times, they refuse, perhaps egged on by major oil producers like Iran and Venezuela who have the most to gain by keeping the world dependent on the fossil fuels they produce and sell to prop up their dictatorships.

A giant poster of Mao looks approvingly at all the CO2 being produced…for Communism.

We need to take immediate action. While we cannot rule out the threat of military force, we should start by using our diplomatic muscle to push for firm action at top-level summits like the Kyoto Protocol. Second, we should fight back against the liberals who are trying to hold up this important work, from big government bureaucrats trying to regulate clean energy to celebrities accusing people who believe in global warming of being 'racist'. Third, we need to continue working with American industries to set an example for the world by decreasing our own emissions in order to protect ourselves and our allies. Finally, we need to punish people and institutions who, instead of cleaning up their own carbon, try to parasitize off the rest of us and expect the federal government to do it for them.

Please join our brave men and women in uniform in pushing for an end to climate change now.

If *this* were the narrative conservatives were seeing on TV and in the papers, I think we'd have action on the climate pretty

quickly. I mean, that action might be nuking China. But it would be action.

And yes, there's a sense in which that narrative is dishonest, or at least has really weird emphases. But our current narrative *also* has really some weird emphases. And for much the same reasons.

VI.

The Red Tribe and Blue Tribe have different narratives, which they use to tie together everything that happens into reasons why their tribe is good and the other tribe is bad.

Sometimes this results in them seizing upon different sides of an apparently nonpolitical issue when these support their narrative; for example, Republicans generally supporting a quarantine against Ebola, Democrats generally opposing it. Other times it results in a side trying to gain publicity for stories that support their narrative while sinking their opponents' preferred stories – Rotherham for some Reds; Ferguson for some Blues.

When an issue gets tied into a political narrative, it stops being about itself and starts being about the wider conflict between tribes until eventually it becomes viewed as a Referendum On Everything. At this point, people who are clued in start suspecting nobody cares about the issue itself – like victims of beheadings, or victims of sexual abuse – and everybody cares about the issue's potential as a political weapon – like proving Muslims are "uncivilized", or proving political correctness is dangerous. After that, even people who agree that the issue is a problem and who would otherwise want to take action have to stay quiet, because they know that their help would be used less to solve a problem than to push forward the war effort against them. If they feel especially threatened, they may even

take an unexpected side on the issue, switching from what they would usually believe to whichever position seems less like a transparent cover for attempts to attack them and their friends.

And then you end up doing silly things like saying ISIS is not as bad as Fox News, or donating hundreds of thousands of dollars to the officer who shot Michael Brown.

This can sort of be prevented by *not* turning everything into a referendum on how great your tribe is and how stupid the opposing tribe is, or by trying to frame an issue in a way that respects or appeals to an out-group's narrative.

Let me give an example. I find a lot of online feminism very triggering, because it seems to me to have nothing to do with women and be transparently about marginalizing nerdy men as creeps who are not really human (see: nude pictures vs. Rotherham, above). This means that even when I support and agree with feminists and want to help them, I am constantly trying to drag my brain out of panic mode that their seemingly valuable projects are just deep cover for attempts to hurt me (see: hypothetical Bill O'Reilly demanding Muslims condemn the "Islamic" practice of beheading people).

I have recently met some other feminists who instead use a narrative which views "nerds" as an "alternative gender performance", ie in the case of men they reject the usual masculine pursuits of sports and fraternities and they have characteristics that violate normative beauty standards (like "no neckbeards"). Thus, people trying to attack nerds is a subcategory of "people trying to enforce gender performance", and nerds should join with queer people, women, and other people who have an interest in promoting tolerance of alternative gender performances in order to fight for their mutual right to be left alone and accepted.

I'm not sure I entirely buy this argument, but it doesn't trigger me, and it's the sort of thing I *could* buy, and if all my friends started saying it I'd probably be roped into agreeing by social pressure alone.

But this is as rare as, well, anti-global warming arguments aimed at making Republicans feel comfortable and nonthreatened.

I blame the media, I really do. Remember, from within a system no one necessarily has an incentive to do what the system as a whole is supposed to do. Daily Kos or someone has a little label saying "supports liberal ideas", but *actually* their incentive is to make liberals want to click on their pages and ads. If the quickest way to do that is by writing story after satisfying story of how dumb Republicans are, and what wonderful taste they have for being members of the Blue Tribe instead of evil mutants, then they'll do that even if the effect on the entire system is to make Republicans hate them and by extension everything they stand for.

I don't know how to fix this.

Black People Less Likely

[Content warning: Polyamory, race]

I.

The best reporting on social science statistics, like the best reporting in most areas, comes from The Onion:

> CAMBRIDGE, MA—A Harvard University study of more than 2,500 middle-income African-American families found that, when compared to other ethnic groups in the same income bracket, blacks were up to 23 percent more likely. "Our data would seem to discredit the notion that black Americans are less likely," said head researcher Russell Waterstone, noting the study also found that women of African descent were no more or less prone than Latinas. "In fact, over the past several decades, we've seen the African-American community nearly triple in probability." The study noted that, furthermore, Asian-Americans.

I thought of this today because a bunch of people have accosted me about the article There's A Big Problem With Polyamory That Nobody's Talking About. "Scott, you're polyamorous! What do you think of this?"

As per the article, the big problem with polyamorous people is:

> …their whiteness. And that standard of whiteness not only erases the experience of people of color; it reflects the actual exclusion of these people in poly life and communities. […]

> A white, affluent image that reflects a troubling reality: A 2013 survey of polyamorous people from online groups, mailing lists and forums found that almost 90% of the participants identified as Caucasian. People of color, especially black polyamorists, report feeling "othered" and excluded in poly environments such as meet-ups, with women feeling especially at risk of being objectified and fetishized as an exotic sexual plaything.
>
> "I interviewed a black couple who went to a poly group, and they were definitely preyed upon, in a sense," said Marla Renee Stewart, Atlanta-based founder of Velvet Lips, a sex education venue.

The article constantly equivocates between "the problem is that polyamory is too white" and "the problem is that the media portrays polyamory as too white", which is kind of a weird combination of problems to be discussing in a media portrayal. But it seems to eventually settle on a thesis that black people really are strongly underrepresented.

For the record, here is a small sample of other communities where black people are strongly underrepresented:

Runners (3%). Bikers (6%). Furries (2%). Wall Street senior management (2%). Occupy Wall Street protesters (unknown but low, one source says 1.6% but likely an underestimate). BDSM (unknown but low) Tea Party members (1%). American Buddhists (~2%). Bird watchers (4%). Environmentalists (various but universally low). Wikipedia contributors (unknown but low). Atheists (2%). Vegetarian activists (maybe 1-5%). Yoga enthusiasts (unknown but low). College baseball players (5%). Swimmers (2%). Fanfiction readers (2%). Unitarian Universalists (1%).

Can you see what all of these groups have in common?

No. No you can't. If there's some hidden factor uniting Wall Street senior management and furries, it is way beyond any of our pay grades.

But what I noticed when I looked up those numbers was that in every case, the people involved have come up with a pat explanation that sounds perfectly plausible right up until you compare it to any other group, at which point it bursts into flames.

For example, Some people explain try to explain declining black interest in baseball by appeal to how some baseball personality made some horribly racist remark. But Donald Sterling continues to be racist as heck, and black people continue to be more than three-quarters of basketball players.

Some people try to explain black people's underrepresentation on fanfiction websites by saying that many of them have limited access to the Internet. Okay. Except that black people are heavily *overrepresented* on Twitter, making up double the expected proportion of that site's population.

Some people try to explain the underrepresentation of blacks in libertarianism and the Tea Party by arguing that these groups' political beliefs are contrary to black people's life experiences. But blacks are also underrepresented in groups with precisely the opposite politics. That they make up only 1.6% of visitors to the Occupy Wall Street website is no doubt confounded by who visits websites, but even people who looked at the protests agree that there was a stunning shortage of black faces. I would have liked to get current membership statistics for the US Communist Party, but they weren't available, so I fudged by looking at the photos of people who "liked" the US Communist Party's Facebook page. 3% of them were black. Blacks are more likely to endorse

environmentalism than whites, but less likely to be involved in the environmentalist movement.

Some people try to explain black people's underrepresentation on Wall Street by saying Wall Street is racist and intolerant. But Unitarian Universalists are just about the most tolerant people in the world – nobody even knows what they do, just that they're extremely tolerant when they do it – and black people are in Unitarianism at lower rates than they're on Wall Street.

And the article on polyamory suggested that maybe polyamorists' high-flying lifestyle and expensive play parties price out black people. Forget for a moment that I've been poly for three years and had no idea this high-flying lifestyle existed and kind of feel like I am missing out. Forget for a moment that as far as I can tell "play parties" are a BDSM term with no relationship to polyamory. In my experience polyamory draws from the same sort of people as atheism, and atheism is *very* white even though not believing in God doesn't cost a cent.

This entire genre seems to be a bunch of really silly ad hoc arguments by people who aren't talking to each other. I would guess most of the underrepresentation of black people in all of these things are for the same couple of reasons.

First, some of these things require some level of affluence – I know I just said that didn't explain polyamory, but I think it explains some others. For example, bird-watching requires you live somewhere suburban or rural where there are interesting birds, want to waste money on binoculars, and have some free time. Swimming requires you live in an area where the schools or at least the neighborhoods have pools.

Second, Maslow's Hierarchy Of Needs says you're not going to do weird things to self-actualize until you feel materially safe and secure. A lot of black people don't feel like they're in a position where they can start worrying about where the best bird-watching is at.

Third, the thrive-survive dichotomy says materially insecure people are going to value community and conformity more. Polyamory is still pretty transgressive, and unless you feel very safe or feel sufficiently mobile and atomized that you don't care what your community thinks about you, you're not going to feel comfortable making that transgression. Many of these things require leaving the general community to participate in a weird insular subculture, and that requires a sort of lack of preexisting community bonds that I think only comes with the upper middle class.

Fourth, black people might avoid weird nonconformist groups because they're already on thin enough ice in terms of social acceptance. Being a black person probably already exposes you to enough stigma, without becoming a furry as well.

Fifth, we already know that neighborhoods and churches tend to end up mostly monoracial through a complicated process of aggregating small acts of self-segregation based on slight preferences not to be completely surrounded by people of a different race. It doesn't seem too unlikely to me that a similar process could act on hobbies and interest groups.

Sixth, even when black people are involved in weird subcultures, they may do them separately from white people, leading white people to think their hobby is almost all white – and leading mostly white academics to miss them in their studies. I once heard about a professor who accused Alcoholics Anonymous of being racist, on the grounds that its

membership was almost entirely white. The (white) professor had surveyed AA groups in his (white) neighborhood and asked his (white) friends and (white) grad students to do the same. Meanwhile, when more sober minds (no pun intended) investigated, they found black areas had thriving majority-black AA communities.

Seventh, a lot of groups are stratified by education level. Black people are only about half as likely to have a bachelor's degree. This matters a lot in areas like atheism that are disproportionately limited to the most educated individuals. Polyamory also falls into this category – the most recent survey found 85% of poly people had a college education, compared to 30% of the general population (!). 30% of poly people had a graduate degree compared to only about 10% of the general population and only about 3% of blacks. There has to be a strong education filter on polyamory to produce those kinds of numbers, and I think that alone is big enough to explain most of the black underrepresentation.

Eighth, people of the same social class tend to cluster, and black people are disproportionately underrepresented among the upper middle class. Most of these fields are dominated by upper middle class people. The nickname for weird self-actualizing upper middle class things is "Stuff White People Like", and this is not a coincidence. [EDIT: Commenter John Schilling says this better than I – a lot of these groups are about differentiating yourself from a presumedly boring low-status middle class existence, but black people fought hard to get into the middle class, or are still fighting, and are less excited about differentiating themselves from it.]

So I think positing that black people feel "fetishized as an exotic sexual plaything" in the poly community is unnecessary. Black people are underrepresented in the poly

community for the same reason they're underrepresented in everything in the same vague circle as poly. Heck, black people are even underrepresented in the activity of complaining about black people being fetishized as exotic sexual playthings – check out Tumblr's racial demographics if you don't believe me.

II.

The eight points above add up to a likelihood that black people will probably be underrepresented in a lot of weird subculturey nonconformist things. This is not a firm law – black people will be overrepresented in a few weird subculturey nonconformist things that are an especially good fit for their culture – but overall I think the rule holds. And that's a big problem.

A few paragraphs back I mentioned that Occupy Wall Street was had disproportionately few minorities. Here are some other people who like to mention this: Michelle Malkin. The Daily Caller. American Thinker. View From The Right. New York Post. American Renaissance.

All of these sources have something in common, and it's *not* a heartfelt concern for equal minority representation.

Likewise, you know who's got an *obsessively* large collection of resources on the underrepresentation of minorities in atheism? Conservapedia (Western Atheism And Race, Racial Demographics Of The Richard Dawkins Audience, Richard Dawkins' Lack Of Appeal To The Asian Woman Audience, etc, etc, not to mention the very classy Richard Dawkins' Family Fortune And The Slave Trade.)

Here it is *easy* to see that "you have low minority representation" serves as a stand-in for "you're racist" serves as a stand-in for "you suck". So here's the problem:

In theocracies ruled by the will of God, people will find that God hates weird people who refuse to conform.

In philosopher-kingdoms ruled by pure reason, people will find that pure reason condemns weird people who refuse to conform.

And in enlightened liberal democracies where we "tolerate anything except intolerance", people will find that weird people who refuse to conform are intolerant.

And if blacks are underrepresented in weird nonconformist groups, and nobody mentions that this is a general principle, that's making their job *way too easy*.

So here's why this article annoys me. In the midst of black underrepresentation in everything in the same ontological category as polyamory, people bring up black underrepresentation in polyamory and suggest it's because poly people are "objectifying" and "preying on" them, positing that "there's a problem" with "a standard of whiteness that erases people of color" in the polyamory community.

We know from OKCupid statistics that (mostly monogamous) white men are very reluctant to date black women, but monogamous people don't have to listen to well-meaning friends going up to them and saying "So, you're mono, I hear the monogamous community has a racism problem."

But now I and other polyamorous people are going to have to answer one more round of annoying questions about "You're polyamorous? Isn't that a bunch of racist nerdy white dudes?"

Nydwracu's Fnords

I.

The fnords first appear in Anton-Wilson and Shea's book *Illuminatus*. Educators, operating as tools of the titular conspiracy, hypnotize all primary school children to have a panic reaction to the trigger word "fnord". The children, who remember nothing of the sessions when they wake up, are incapable of registering the word except as an unexplained feeling of unease.

This turns them into helpless, easily herded adults. Every organ of the media – newspapers, books, cable TV – contains a greater or lesser number of fnords. When some information is counter to the aims of the conspiracy – maybe a communist party organizing in a state where the conspiracy wears a capitalist hat – the secret masters don't bother censoring or suppressing it. Instead, the newspaper reports it on the front page, but fills the article with fnords. Most people read partway through, become very uncomfortable and upset without knowing why, and decide that communists are *definitely* bad people for some reason or other and there's no reason they need to continue reading the article. Why should they worry about awful things like that when there's the whole rest of the paper to read?

According to the book, the only section of the newspaper without any fnords at all is the advertisements.

II.

Last week, some Internet magazine published the latest attempt at the genre of Did You Know Neoreaction Exists You Should Be Outraged. A couple of reactionaries wrote the usual

boring “actually, nothing you said was true, why would you say false things?” responses. Nydwracu, a frequent commenter on this blog, did something I thought was *much* more interesting. He wrote a post called Fnords where he removed all of the filler words and transitions between ideas and thin veneer of argument until he stripped the essay down to the bare essentials. It looked like this:

> Mouthbreathing Machiavellis Dream Of A Silicon Reich strange and ultimately doomed stunt flamboyant act of corporate kiss-assery latest political fashion California Confederacy total corporate despotism potent bitter Steve Jobs Ayn Rand Ray Kurzweil prominent divisive fixture hard-right seditionist aggressively dogmatic blogger reverent following in certain tech circles prolific incomprehensible vanguard youngish white males embittered by “political correctness” Blade Runner, but without all those Asian people cluttering up the streets like to see themselves as the heroes of another sci-fi movie “redpilled” The Matrix “genius” a troll who belches from the depths of an Internet rabbit hole frustrated poet cranky letters to alternative weekly newspapers preoccupations with domineering strongmen angry pseudonym J.R.R. Tolkien George Lucas typical keyboard kook archaic, grandiose snippets cherry-picked from obscure old lack of higher ed creds overconfident autodidact’s imitation fascist teenage Dungeon Master most toxic arguments snugly wrapped in purple prose and coded language oppressive nexus teeth-gnashing white supremacists who haunt the web “men’s rights” advocates nuts disillusioned typical smarmy, meandering (Sure. Easy!) Incredible as it sounds, absolute dictatorship may be the least objectionable tenet espoused by the Dark

> Enlightenment neoreactionaries. Chinese eugenics impending global reign of “autistic nerds These imaginary übermensch sprawling network of blogs, sub-Reddits old-timey tyrants basically racism scientific-sounding euphemism familiar tropes of white victimhood perhaps best known for his infamous slavery apologia poor, persecuted Senator Joe McCarthy. Big surprise. pseudo-intellectual equivalent of a Gwar concert, one sick stunt after another, calculated to shock the attention he so transparently craves “silly not scary” “all of these people need to relax: P.G. Wodehouse football get drunk Internet curio “sophisticated neo-fascism” must be confronted “creepy” future-fascist dictator sadly Koch brothers no matter how crazy your ideas are, radicalism neoreactionaries flatter the prejudices of the new Silicon Valley elite enemies patchwork map of feudal Europe Forget universal rights; signposts of the neoreactionary fantasyland anti-democratic authoritarianism bigotry blue-sea libertarian dream extreme libertarian advocacy Ted Cruz libertarian a small and shallow world a dictatorial approach mythical “god-kings” Stupid proles! They don’t deserve our brilliance! shockingly common would never occur to other people precisely because they’ve refused to leave that stage of youthful live forever escape to outer space or an oceanic city-state play chess against a robot that can discuss Tolkien fantasies childhood imagination perhaps too generous the fundamental problem with these mouthbreathers’ dreams of monarchy. They’ve never role-played the part of the peasant.

That…sure gives one a different perspective on political discourse. I am reminded of those Renaissance artists who

secretly cut up cadavers to learn what was inside people, and from then on all of their human figures would be a little bit creepy because you could almost *see* how the internal bones and muscles were animating the flesh.

Since no one is meta and everyone only pays attention to things when it's their own opinions under threat, I suppose I have to do the same thing with an article from some website on the right:

> socialism completely government run pure single-payer "an island of socialism in American healthcare" that won't change a thing in fact it's a distraction excessive delays tragically predictable bureaucratic rationing price controls, inefficiencies, and the inevitable cover-ups bureaucratic incentives statist VA healthcare system mirrors the government-run healthcare problems slip-shod failure run-amok bureaucrats don't tell me the problem is not enough government money the Paul Krugmans of the world and their leftist allies socialist medicine socialism doesn't work who opposed market choice and competition Senator Harry Reid and House Democratic leader Nancy Pelosi Obamacare job-destroying tax and regulatory provisions

Interestingly, both of those came out to between 13 and 14% of the length of the original article. I wonder if that's some kind of iron law.

III.

I don't know if he ever read *Illuminatus* or whether it was just one of those coincidences, but Jonathan Haidt did the thing with the fnords in real life.

(Warning: a tangentially related study by the same group has recently failed to replicate)

He wanted to test the role of disgust in moral judgments. So he hypnotized a bunch of people to feel disgust at a trigger word – "takes" for half the participants, "often" for the other half – and hypnotically instructed them to forget all about this. Then in an "unrelated study" he asked them to rate the morality of different ethically controversial vignettes. For example:

"A brother and sister fall in love with each other. They frequently **take** vacations together where they have sex. Both are freely consenting and she is on very careful birth control."

or

"A brother and sister fall in love with each other. They **often** go on vacations together where they have sex. Both are freely consenting and she is on very careful birth control."

The participants hypnotized to hate the word "take" found the behavior more objectionable with the "take" version of the vignette than the "often" version, and the participants hypnotized to hate the word "often" displayed the opposite pattern. When they asked subjects to explain their judgment, they gave perfectly reasonable explanations, which could be anything from "incest is just wrong" to "what if they have a child and it's deformed, yeah, I know it said they were on birth control, but it still bothers me."

Then Haidt and his team presented the following story:

"Dan is student council president. It is his job to pick topics for discussion at student meetings. He frequently **takes** suggestions from students and teachers on which topic to choose."

or

“Dan is student council president. It is his job to pick topics for discussion at student meetings. He **often** accepts suggestions from students and teachers on which topic to choose.”

Participants were asked to judge how evil a person Dan was. And when their trigger word was in the sentence, their answer was: pretty evil! When asked to explain themselves, they came up with weird justifications like “Dan is a popularity-seeking snob” or “It just seems he’s up to something”

IV.

A few weeks ago, I noticed something strange.

Every time someone complaints about climate denial, they make extraordinary efforts to get the name of the Koch brothers in. Like it’s never just “Why do so many people believe climate denialism?” it’s more “Why do so many people believe climate denialism, as funded by people like the Koch brothers?”

This is strange because it seems to me that they are acting like associating climate denialism with the Koch brothers will lower its credibility or make it sound vaguely evil.

But this shouldn’t work. The only thing the average person knows about the Koch brothers is that they are people who fund climate change denial. So if you already don’t like climate change denial, this will make you dislike the Koch brothers. But mentioning “Koch brothers!” won’t make you dislike climate change denial more, it will just remind you of one of the downstream effects of your disliking climate change (not liking the Kochs). On the other hand, if you’re still neutral on climate change denial, then you have no reason to dislike the Kochs, and mentioning them won’t help you there either. And if you actively support climate denial, you probably think

the Koch brothers are heroes, so associating them with the movement won't be a good way of discrediting it.

Basically, since your opinion of the Koch brothers should equal your opinion of climate denial, trying to tar climate denial by association with the Kochs is trying to make people dislike an idea by linking it to itself. It shouldn't work.

But I think it does. When you read articles on the other side, they always mention Al Gore. In fact, there are a lot of these people who get brought up as bogeymen every so often.

I have two boring hypotheses and an interesting one.

The first boring hypothesis is that the Koch brothers are white male billionaires. This is enough to make them suspicious. Therefore, global warming skepticism is tarred by association with them, even though we know nothing else about them.

The second boring hypothesis is that it doesn't matter who the Koch brothers are, what matters is the claim that there is some figure funding the movement, that it's not a grassroots upswelling of people genuinely doubtful of global warming, but just one guy (well, two guys) trying to inflict their own weird contrarianism on everyone else.

The interesting hypothesis is that the brain is going loopy, having one of those rare experiences where it forgets not to condition on itself.

Imagine that you don't like climate denialism. You hear that the Koch brothers support climate denialism. You use that information to decide you don't like the Koch brothers very much.

Then a month passes and you forget *exactly* why you don't like the Koch brothers. You just have a very strong feeling that "it just seems like they're up to something."

Then someone tells you the Koch brothers support denialism. And you say: "If *those* bastards support it, then I hate it *even more*!"

In other words, you have undergone a two step process to ratchet up your dislike of climate denialism by associating it with itself.

> We know this idea is evil because it's pushed by such terrible people. We know the people are terrible because they push such an evil idea.
>
> — Scott Alexander (@slatestarcodex) May 18, 2014

I wonder if this is part of what makes politics so divisive. You start off with a weak preference in one direction. Gradually, certain words like "Koch brothers" or "Exxon Mobil" become fnords, reservoirs of your negative feelings, and then every time you read about climate change, even if there's no real argument, you get triggered and become pretty sure denialists are up to something, in the same way Dan the student council president is up to something. And the other side gets different fnords – "Climategate", "hockey stick graph", and they go through the same process. And finally you get totally incomprehensible arguments: "But how can you be a climate change denier when that associates you with the Koch brothers?! Did you know climate change denialism is *literally* sponsored by the Heartland Institute?!" And the other side is just nodding their head and going "Oh, yeah, my sister used to work there."

V.

IF YOU DON'T SEE THE FNORD IT CAN'T EAT YOU

All in All, Another Brick in the Motte

One of the better things I've done with this blog was help popularize Nicholas Shackel's "motte and bailey doctrine". But I've recently been reminded I didn't do a very good job of it. The original discussion is in the middle of a post so controversial that it probably can't be linked in polite company – somewhat dampening its ability to popularize anything.

In order to rectify the error, here is a nice clean post on the concept that adds a couple of further thoughts to the original formulation.

The original Shackel paper is intended as a critique of post-modernism. Post-modernists sometimes say things like "reality is socially constructed", and there's an uncontroversially correct meaning there. We don't experience the world directly, but through the categories and prejudices implicit to our society; for example, I might view a certain shade of bluish-green as blue, and someone raised in a different culture might view it as green. Okay.

Then post-modernists go on to say that if someone in a different culture thinks that the sun is light glinting off the horns of the Sky Ox, that's just as real as our own culture's theory that the sun is a mass of incandescent gas a great big nuclear furnace. If you challenge them, they'll say that you're denying reality is socially constructed, which means you're clearly very naive and think you have perfect objectivity and the senses perceive reality directly.

The writers of the paper compare this to a form of medieval castle, where there would be a field of desirable and economically productive land called a bailey, and a big ugly

tower in the middle called the motte. If you were a medieval lord, you would do most of your economic activity in the bailey and get rich. If an enemy approached, you would retreat to the motte and rain down arrows on the enemy until they gave up and went away. Then you would go back to the bailey, which is the place you wanted to be all along.

So the motte-and-bailey doctrine is when you make a bold, controversial statement. Then when somebody challenges you, you claim you were just making an obvious, uncontroversial statement, so you are clearly right and they are silly for challenging you. Then when the argument is over you go back to making the bold, controversial statement.

Some classic examples:

1. The religious group that acts for all the world like God is a supernatural creator who builds universes, creates people out of other people's ribs, parts seas, and heals the sick when asked very nicely (bailey). Then when atheists come around and say maybe there's no God, the religious group objects "But God is just another name for the beauty and order in the Universe! You're not denying that there's beauty and order in the Universe, are you?" (motte). Then when the atheists go away they get back to making people out of other people's ribs and stuff.

2. Or…"If you don't accept Jesus, you will burn in Hell forever." (bailey) But isn't that horrible and inhuman? "Well, Hell is just another word for being without God, and if you choose to be without God, God will be nice and let you make that choice." (motte) Oh, well that doesn't sound so bad, I'm going to keep rejecting Jesus. "But if you reject Jesus, you will BURN in HELL FOREVER and your body will be GNAWED

BY WORMS." But didn't you just… "Metaphorical worms of godlessness!"

3. The feminists who constantly argue about whether you can be a real feminist or not without believing in X, Y and Z and wanting to empower women in some very specific way, and who demand everybody support controversial policies like affirmative action or affirmative consent laws (bailey). Then when someone says they don't really like feminism very much, they object "But feminism is just the belief that women are people!" (motte) Then once the person hastily retreats and promises he *definitely* didn't mean women aren't people, the feminists get back to demanding everyone support affirmative action because feminism, or arguing about whether you can be a feminist and wear lipstick.

4. Proponents of pseudoscience sometimes argue that their particular form of quackery will cure cancer or take away your pains or heal your crippling injuries (bailey). When confronted with evidence that it doesn't work, they might argue that people need hope, and even a placebo solution will often relieve stress and help people feel cared for (motte). In fact, some have argued that quackery may be better than real medicine for certain untreatable diseases, because neither real nor fake medicine will help, but fake medicine tends to be more calming and has fewer side effects. But then once you leave the quacks in peace, they will go back to telling less knowledgeable patients that their treatments will cure cancer.

5. Critics of the rationalist community note that it pushes controversial complicated things like Bayesian statistics and utilitarianism (bailey) under the name "rationality", but when asked to justify itself defines rationality as "whatever helps you achieve your goals", which is so vague as to be universally unobjectionable (motte). Then once you have

admitted that more rationality is always a good thing, they suggest you've admitted everyone needs to learn more Bayesian statistics.

6. Likewise, singularitarians who predict with certainty that there will be a singularity, because "singularity" just means "a time when technology is so different that it is impossible to imagine" – and really, who would deny that technology will probably get really weird (motte)? But then every other time they use "singularity", they use it to refer to a very specific scenario of intelligence explosion, which is far less certain and needs a lot more evidence before you can predict it (bailey).

The motte and bailey doctrine sounds kind of stupid and hard-to-fall-for when you put it like that, but *all* fallacies sound that way *when you're thinking about them*. More important, it draws its strength from people's usual failure to debate specific propositions rather than vague clouds of ideas. If I'm debating "does quackery cure cancer?", it might be easy to view that as a general case of the problem of "is quackery okay?" or "should quackery be illegal?", and from there it's easy to bring up the motte objection.

Recently, a friend (I think it was Robby Bensinger) pointed out something I'd totally missed. The motte-and-bailey doctrine is a perfect mirror image of my other favorite fallacy, the weak man fallacy.

Weak-manning is a lot like straw-manning, except that instead of debating a fake, implausibly stupid opponent, you're debating a real, unrepresentatively stupid opponent. For example, "Religious people say that you should kill all gays. But this is evil. Therefore, religion is wrong and barbaric. Therefore we should all be atheists." There are certainly religious people who think that you should kill all gays, but

they're a small fraction of all religious people and probably not the ones an unbiased observer would hold up as the best that religion has to offer.

If you're debating the Pope or something, then when you weak-man, you're unfairly replacing a strong position (the Pope's) with a weak position (that of the guy who wants to kill gays) to make it more attackable.

But in motte and bailey, you're unfairly replacing a weak position (there is a supernatural creator who can make people out of ribs) with a strong position (there is order and beauty in the universe) in order to make it more defensible.

So weak-manning is replacing a strong position with a weak position to better attack it; motte-and-bailey is replacing a weak position with a strong position to better defend it.

This means people who know both terms are at constant risk of arguments of the form "You're weak-manning me!" "No, *you're* motte-and-baileying *me!*".

Suppose we're debating feminism, and I defend it by saying it really is important that women are people, and you attack it by saying that it's not true that all men are terrible. Then I can accuse you of making life easy for yourself by attacking the weakest statement anyone vaguely associated with feminism has ever pushed. And you can accuse me if making life too easy for myself by defending the most uncontroversially obvious statement I can get away with.

So what is the *real* feminism we should be debating? *Why would you even ask that question?* What is this, some kind of dumb high school debate club? Who the heck thinks it would be a good idea to say "Here's a vague poorly-defined concept that mind-kills everyone who touches it – quick, should you associate it with positive affect or negative affect?!"

Taboo your words, then replace the symbol with the substance. If you have an *actual thing* you're trying to debate, then it should be obvious when somebody's changing the topic. If working out who's using motte-and-bailey (or weak man) is remotely difficult, it means your discussion went wrong several steps earlier and you probably have no idea what you're even arguing about.

PS: Nicholas Shackel, original inventor of the term, weighs in.

Ethnic Tension and Meaningless Arguments

I.

Part of what bothers me – and apparently several others – about yesterday's motte-and-bailey discussion is that here's a fallacy – a pretty successful fallacy – that depends entirely on people not being entirely clear on what they're arguing about. Somebody says God doesn't exist. Another person objects that God is just a name for the order and beauty in the universe. Then this somehow helps defend the position that God is a supernatural creator being. How does that even happen?

"Sir, you've been accused of murdering your wife. We have three witnesses who said you did it. What do you have to say for yourself?"

"Well, your honor, I think it's quite clear I didn't murder the President. For one thing, he's surrounded by Secret Service agents. For another, check the news. The President's still alive."

"Huh. For some reason I vaguely remember thinking you didn't have a case. Yet now that I hear you talk, everything you say is incredibly persuasive. You're free to go."

While motte-and-bailey is less subtle, it seems to require a similar sort of misdirection. I'm not saying it's impossible. I'm just saying it's a fact that needs to be explained.

When everything works the way it's supposed to in philosophy textbooks, arguments are supposed to go one of a couple of ways:

1. Questions of empirical fact, like “Is the Earth getting warmer?” or “Did aliens build the pyramids?”. You debate these by presenting factual evidence, like “An average of global weather station measurements show 2014 is the hottest year on record” or “One of the bricks at Giza says ‘Made In Tau Ceti V’ on the bottom.” Then people try to refute these facts or present facts of their own.

2. Questions of morality, like “Is it wrong to abort children?” or “Should you refrain from downloading music you have not paid for?” You can only debate these *well* if you’ve already agreed upon a moral framework, like a particular version of natural law or consequentialism. But you can *sort of* debate them by comparing to examples of agreed-upon moral questions and trying to maintain consistency. For exmaple, “You wouldn’t kill a one day old baby, so how is a nine month old fetus different?” or “You wouldn’t download a *car*.”

If you are very lucky, your philosophy textbook will also admit the existence of:

3. Questions of policy, like “We should raise the minimum wage” or “We should bomb Foreignistan”. These are combinations of competing factual claims and competing values. For example, the minimum wage might hinge on factual claims like “Raising the minimum wage would increase unemployment” or “It is very difficult to live on the minimum wage nowadays, and many poor families cannot afford food.” But it might also hinge on value claims like “Corporations owe it to their workers to pay a living wage,” or “It is more important that the poorest be protected than that the economy be strong.” Bombing Foreignistan might depend on factual claims like “The Foreignistanis are harboring terrorists”, and on value claims like “The safety of our people is worth the risk of collateral damage.” If you can resolve all

of these factual and value claims, you should be able to agree on questions of policy.

None of these seem to allow the sort of vagueness of topic mentioned above.

II.

A question: are you pro-Israel or pro-Palestine? Take a second, actually think about it.

Some people probably answered pro-Israel. Other people probably answered pro-Palestine. Other people probably said they were neutral because it's a complicated issue with good points on both sides.

Probably very few people answered: *Huh? What?*

This question doesn't fall into any of the three Philosophy 101 forms of argument. It's not a question of fact. It's not a question of particular moral truths. It's not even a question of policy. There are closely related policies, like whether Palestine should be granted independence. But if I support a very specific two-state solution where the border is drawn upon the somethingth parallel, does that make me pro-Israel or pro-Palestine? At exactly which parallel of border does the solution under consideration switch from pro-Israeli to pro-Palestinian? Do you think the crowd of people shouting and waving signs saying "SOLIDARITY WITH PALESTINE" have an answer to that question?

But it's even worse, because this question covers much more than just the borders of an independent Palestinian state. Was Israel justified by responding to Hamas' rocket fire by bombing Gaza, even with the near-certainty of collateral damage? Was Israel justified in building a wall across the Palestinian territories to protect itself from potential terrorists,

even though it severely curtails Palestinian freedom of movement? Do Palestinians have a "right of return" to territories taken in the 1948 war? Who should control the Temple Mount?

These are four very different questions which one would think each deserve independent consideration. But in reality, what percent of the variance in people's responses do you think is explained by a general "pro-Palestine vs. pro-Israel" factor? 50%? 75%? More?

In a way, when we round people off to the Philosophy 101 kind of arguments, we are failing to respect their self-description. People aren't out on the streets saying "By my cost-benefit analysis, Israel was in the right to invade Gaza, although it may be in the wrong on many of its other actions." They're waving little Israeli flags and holding up signs saying "ISRAEL: OUR STAUNCHEST ALLY". Maybe we should take them at face value.

This is starting to look related to the original question in (I). Why is it okay to suddenly switch points in the middle of an argument? In the case of Israel and Palestine, it might be because people's support for any particular Israeli policy is better explained by a General Factor Of Pro-Israeliness than by the policy itself. As long as I'm arguing in favor of Israel in *some way*, it's still considered by everyone to be on topic.

III.

Some moral philosophers got fed up with nobody being able to explain what the heck a moral truth was and invented emotivism. Emotivism says there *are* no moral truths, just expressions of little personal bursts of emotion. When you say "Donating to charity is good," you don't mean "Donating to charity increases the sum total of utility in the world," or

“Donating to charity is in keeping with the Platonic moral law” or “Donating to charity was commanded by God” or even “I like donating to charity”. You’re just saying “Yay charity!” and waving a little flag.

Seems a lot like how people handle the Israel question. “I’m pro-Israel” doesn’t necessarily imply that you believe any empirical truths about Israel, or believe any moral principles about Israel, or even support any Israeli policies. It means you’re waving a little flag with a Star of David on it and cheering.

So here is Ethnic Tension: A Game For Two Players.

Pick a vague concept. “Israel” will do nicely for now.

Player 1 tries to associate the concept “Israel” with as much good karma as she possibly can. Concepts get good karma by doing good moral things, by being associated with good people, by being linked to the beloved in-group, and by being oppressed underdogs in bravery debates.

“Israel is the freest and most democratic country in the Middle East. It is one of America’s strongest allies and shares our Judeo-Christian values.

Player 2 tries to associate the concept “Israel” with as much bad karma as she possibly can. Concepts get bad karma by committing atrocities, being associated with bad people, being linked to the hated out-group, and by being oppressive big-shots in bravery debates. Also, she obviously needs to neutralize Player 1’s actions by disproving all of her arguments.

“Israel may have some level of freedom for its most privileged citizens, but what about the millions of people in the Occupied Territories that have no say? Israel is involved in various

atrocities and has often killed innocent protesters. They are essentially a neocolonialist state and have allied with other neocolonialist states like South Africa."

The prize for winning this game is the ability to win the other three types of arguments. If Player 1 wins, the audience ends up with a strongly positive General Factor Of Pro-Israeliness, and vice versa.

Remember, people's capacity for motivated reasoning is pretty much infinite. Remember, a motivated skeptic asks if the evidence *compels* them to accept the conclusion; a motivated credulist asks if the evidence *allows* them to accept the conclusion. Remember, Jonathan Haidt and his team hypnotized people to have strong disgust reactions to the word "often", and then tried to hold in their laughter when people in the lab came up with convoluted yet plausible-sounding arguments against any policy they proposed that included the word "often" in the description.

I've never heard of the experiment being done the opposite way, but it sounds like the sort of thing that might work. Hypnotize someone to have a very positive reaction to the word "often" (for most hilarious results, have it give people an orgasm). "Do you think governments should raise taxes more often?" "Yes. Yes yes YES YES OH GOD YES!"

Once you finish the Ethnic Tension Game, you're replicating Haidt's experiment with the word "Israel" instead of the word "often". Win the game, and any pro-Israel policy you propose will get a burst of positive feelings and tempt people to try to find some explanation, any explanation, that will justify it, whether it's invading Gaza or building a wall or controlling the Temple Mount.

So this is the fourth type of argument, the kind that doesn't make it into Philosophy 101 books. The trope namer is Ethnic Tension, but it applies to anything that can be identified as a Vague Concept, or paired opposing Vague Concepts, which you can use emotivist thinking to load with good or bad karma.

IV.

Now motte-and-bailey stands revealed:

> Somebody says God doesn't exist. Another person objects that God is just a name for the order and beauty in the universe. Then this somehow helps defend the position that God is a supernatural creator being. How does that even happen?

The two-step works like this. First, load "religion" up with good karma by pitching it as persuasively as possible. "Religion is just the belief that there's beauty and order in the universe."

Wait, *I* think there's beauty and order in the universe!

"Then you're religious too. We're all religious, in the end, because religion is about the common values of humanity and meaning and compassion sacrifice beauty of a sunrise Gandhi Buddha Sufis St. Francis awe complexity humility wonder Tibet the Golden Rule love."

Then, once somebody has a strongly positive General Factor Of Religion, it doesn't really matter whether someone believes in a creator God or not. If they have any predisposition whatsoever to do so, they'll find a reason to let themselves. If they can't manage it, they'll say it's true "metaphorically" and continue to act upon every corollary of it being true.

("God is just another name for the beauty and order in the universe. But Israel definitely belongs to the Jews, because the beauty and order of the universe promised it to them.")

If you're an atheist, you probably have a lot of important issues on which you want people to consider non-religious answers and policies. And if somebody can maintain good karma around the "religion" concept by believing God is the order and beauty in the universe, then that can still be a victory for religion even if it is done by jettisoning many traditionally "religious" beliefs. In this case, it is useful to think of the "order and beauty" formulation as a "motte" for the "supernatural creator" formulation, since it's allowing the *entire concept* to be defended.

But even this is giving people too much credit, because the existence of God is a (sort of) factual question. From yesterday's post:

> Suppose we're debating feminism, and I defend it by saying it really is important that women are people, and you attack it by saying that it's not true that all men are terrible. What is the real feminism we should be debating? Why would you even ask that question? What is this, some kind of dumb high school debate club? Who the heck thinks it would be a good idea to say 'Here's a vague poorly-defined concept that mind-kills everyone who touches it – quick, should you associate it with positive affect or negative affect?!'

Who the heck thinks that? Everybody, all the time.

Once again, if I can load the concept of "feminism" with good karma by making it so obvious nobody can disagree with it, then I have a massive "home field advantage" when I'm trying

to convince anyone of any particular policy that can go under the name "feminism", even if it's unrelated to the arguments that gave feminism good karma in the first place.

Or if I'm against feminism, I just post quotes from the ten worst feminists on Tumblr again and again until the entire movement seems ridiculous and evil, and then you'll have trouble convincing anyone of *anything* feminist. "That seems reasonable…but wait, isn't that a feminist position? Aren't those the people I hate?"

(compare: most Americans oppose Obamacare, but most Americans support each individual component of Obamacare when it is explained without using the word "Obamacare")

V.

Little flow diagram things make everything better. Let's make a little flow diagram thing.

We have our node "Israel", which has either good or bad karma. Then there's another node close by marked "Palestine". We would expect these two nodes to be pretty anti-correlated. When Israel has strong good karma, Palestine has strong bad karma, and vice versa.

Now suppose you listen to Noam Chomsky talk about how strongly he supports the Palestinian cause and how much he dislikes Israel. One of two things can happen:

"Wow, a great man such as Noam Chomsky supports the Palestinians! They must be very deserving of support indeed!"

or

"That idiot Chomsky supports Palestine? Well, screw him. And screw them!"

So now there is a third node, Noam Chomsky, that connects to both Israel and Palestine, and we have discovered it is positively correlated with Palestine and negatively correlated with Israel. It probably has a pretty low weight, because there are a lot of reasons to care about Israel and Palestine other than Chomsky, and a lot of reasons to care about Chomsky other than Israel and Palestine, but the connection is there.

I don't know anything about neural nets, so maybe this system isn't actually a neural net, but whatever it is I'm thinking of, it's a structure where eventually the three nodes reach some kind of equilibrium. If we start with someone liking Israel and Chomsky, but not Palestine, then either that's going to shift a little bit towards liking Palestine, or shift a little bit towards disliking Chomsky.

Now we add more nodes. Cuba seems to really support Palestine, so they get a positive connection with a little bit of weight there. And I think Noam Chomsky supports Cuba, so we'll add a connection there as well. Cuba is socialist, and that's one of the most salient facts about it, so there's a heavily weighted positive connection between Cuba and socialism. Palestine kind of makes noises about socialism but I don't think they have any particular economic policy, so let's say very weak direct connection. And Che is heavily associated with Cuba, so you get a pretty big Che – Cuba connection, plus a strong direct Che – socialism one. And those pro-Palestinian students who threw rotten fruit at an Israeli speaker also get a little path connecting them to "Palestine" – hey, why not – so that if you support Palestine you might be willing to excuse what they did and if you oppose them you might be a little less likely to support Palestine.

Back up. This model produces crazy results, like that people who like Che are more likely to oppose Israel bombing Gaza.

That's such a weird, implausible connection that it casts doubt upon the entire…

Oh. Wait. Yeah. Okay.

I think this kind of model, in its efforts to sort itself out into a ground state, might settle on some kind of General Factor Of Politics, which would probably correspond pretty well to the left-right axis.

In Five Case Studies On Politicization, I noted how fresh new unpoliticized issues, like the Ebola epidemic, were gradually politicized by connecting them to other ideas that were already part of a political narrative. For example, a quarantine against Ebola would require closing the borders. So now there's a weak negative link between "Ebola quarantine" and "open borders". If your "open borders" node has good karma, now you're a little less likely to support an Ebola quarantine. If "open borders" has bad karma, a little more likely.

I also tried to point out how you could make different groups support different things by changing your narrative a little:

> Global warming has gotten inextricably tied up in the Blue Tribe narrative: Global warming proves that unrestrained capitalism is destroying the planet. Global warming disproportionately affects poor countries and minorities. Global warming could have been prevented with multilateral action, but we were too dumb to participate because of stupid American cowboy diplomacy. Global warming is an important cause that activists and NGOs should be lauded for highlighting. Global warming shows that Republicans are science denialists and probably all creationists. Two lousy sentences on "patriotism" aren't going to break through that.

If I were in charge of convincing the Red Tribe to line up behind fighting global warming, here's what I'd say:

In the 1950s, brave American scientists shunned by the climate establishment of the day discovered that the Earth was warming as a result of greenhouse gas emissions, leading to potentially devastating natural disasters that could destroy American agriculture and flood American cities. As a result, the country mobilized against the threat. Strong government action by the Bush administration outlawed the worst of these gases, and brilliant entrepreneurs were able to discover and manufacture new cleaner energy sources. As a result of these brave decisions, our emissions stabilized and are currently declining.

Unfortunately, even as we do our part, the authoritarian governments of Russia and China continue to industralize and militarize rapidly as part of their bid to challenge American supremacy. As a result, Communist China is now by far the world's largest greenhouse gas producer, with the Russians close behind. Many analysts believe Putin secretly welcomes global warming as a way to gain access to frozen Siberian resources and weaken the more temperate United States at the same time. These countries blow off huge disgusting globs of toxic gas, which effortlessly cross American borders and disrupt the climate of the United States. Although we have asked them to stop several times, they refuse, perhaps egged on by major oil producers like Iran and Venezuela who have the most to gain by keeping the world dependent on the fossil fuels they produce and sell to prop up their dictatorships.

> We need to take immediate action. While we cannot rule out the threat of military force, we should start by using our diplomatic muscle to push for firm action at top-level summits like the Kyoto Protocol. Second, we should fight back against the liberals who are trying to hold up this important work, from big government bureaucrats trying to regulate clean energy to celebrities accusing people who believe in global warming of being 'racist'. Third, we need to continue working with American industries to set an example for the world by decreasing our own emissions in order to protect ourselves and our allies. Finally, we need to punish people and institutions who, instead of cleaning up their own carbon, try to parasitize off the rest of us and expect the federal government to do it for them.

In the first paragraph, "global warming" gets positively connected to concepts like "poor people and minorities" and "activists and NGOs", and gets negatively connected to concepts like "capitalism", "American cowboy diplomacy", and "creationists". That gives global warming really strong good karma if (and only if) you like the first two concepts and hate the last three.

In the next three paragraphs, "global warming" gets positively connected to "America", "the Bush administration" and "entrepreneurs", and negatively connected to "Russia", "China", "oil producing dictatorships like Iran and Venezuela", "big government bureaucrats", and "welfare parasites". This is going to appeal to, well, a different group.

Notice two things here. First, the exact connection isn't that important, as long as we can hammer in the existence of a connection. I could probably just say GLOBAL WARMING!

COMMUNISM! GLOBAL WARMING! COMMUNISM! GLOBAL WARMING! COMMUNISM! several hundred times and have the same effect if I could get away with it (this is the principle behind attack ads which link a politician's face to scary music and a very concerned voice).

Second, there is no attempt whatsoever to challenge the idea that the issue at hand is the positive or negative valence of a concept called "global warming". At no point is it debated what the solution is, which countries the burden is going to fall on, or whether any particular level of emission cuts would do more harm than good. It's just accepted as obvious by both sides that we debate "for" or "against" global warming, and if the "for" side wins then they get to choose some solution or other or whatever oh god that's so boring can we get back to Israel vs. Palestine.

Some of the scientists working on IQ have started talking about "hierarchical factors", meaning that there's a general factor of geometry intelligence partially correlated with other things into a general factor of mathematical intelligence partially correlated with other things into a general factor of total intelligence.

I would expect these sorts of things to work the same way. There's a General Factor Of Global Warming that affects attitudes toward pretty much all proposed global warming solutions, which is very highly correlated with a lot of other things to make a General Factor Of Environmentalism, which itself is moderately highly correlated with other things into the General Factor Of Politics.

VI.

Speaking of politics, a fruitful digression: what the heck was up with the Ashley Todd mugging hoax in 2008?

Back in the 2008 election, a McCain campaigner claimed (falsely, it would later turn out) to have been assaulted by an Obama supporter. She said he slashed a "B" (for "Barack") on her face with a knife. This got a lot of coverage, and according to Wikipedia:

> John Moody, executive vice president at Fox News, commented in a blog on the network's website that "this incident could become a watershed event in the 11 days before the election," but also warned that "if the incident turns out to be a hoax, Senator McCain's quest for the presidency is over, forever linked to race-baiting."

Wait. One Democrat, presumably not acting on Obama's direct orders, attacks a Republican woman. And this is supposed to *alter the outcome of the entire election*? In what universe does one crime by a deranged psychopath change whether Obama's tax policy or job policy or bombing-scary-foreigners policy is better or worse than McCain's?

Even *if* we're willing to make the irresponsible leap from "Obama is supported by psychopaths, therefore he's probably a bad guy," there are like a hundred million people on each side. Psychopaths are usually estimated at about 1% of the population, so any movement with a million people will already have 10,000 psychopaths. Proving the existence of a single one changes *nothing*.

I think insofar as this affected the election – and everyone seems to have agreed that it might have – it hit President Obama with a burst of bad karma. Obama something something psychopath with a knife. Regardless of the exact content of those something somethings, *is that the kind of guy you want to vote for*?

Then when it was discovered to be a hoax, it was McCain something something race-baiting hoaxer. Now *he's* got the bad karma!

This sort of conflation between a cause and its supporters really only makes sense in the emotivist model of arguing. I mean, this shouldn't even get dignified with the name *ad hominem* fallacy. Ad hominem fallacy is "McCain had sex with a goat, therefore whatever he says about taxes is invalid." At least it's still the same *guy*. This is something the philosophy textbooks can't bring themselves to believe really exists, even as a fallacy.

But if there's a General Factor Of McCain, then anything bad remotely connected to the guy – goat sex, lying campaigners, whatever – reflects on everything else about him.

This is the same pattern we see in Israel and Palestine. How many times have you seen a news story like this one: "Israeli speaker hounded off college campus by pro-Palestinian partisans throwing fruit. Look at the intellectual bankruptcy of the pro-Palestinian cause!" It's clearly intended as an argument for *something* other than just not throwing fruit at people. The causation seems to go something like "These particular partisans are violating the usual norms of civil discussion, therefore they are bad, therefore something associated with Palestine is bad, therefore your General Factor of Pro-Israeliness should become more strongly positive, therefore it's okay for Israel to bomb Gaza." Not usually said in those *exact words*, but the thread can be traced.

VII.

Here is a prediction of this model: we will be obsessed with what concepts we can connect to other concepts, even when the connection is totally meaningless.

Suppose I say: “Opposing Israel is anti-Semitic”. Why? Well, the Israelis are mostly Jews, so in a sense by definition being anti- them is “anti-Semitic”, broadly defined. Also, p(opposes Israel|is anti-Semitic) is probably pretty high, which sort of lends some naive plausibility to the idea that p(is anti-Semitic|opposes Israel) is at least higher than it otherwise *could* be.

Maybe we do our research and we find exactly what percent of opponents of Israel endorse various anti-Semitic statements like “I hate all Jews” or “Hitler had some bright ideas”. We’ve replaced the symbol with the substance. Problem solved, right?

Maybe not. In the same sense that people can agree on all of the characteristics of Pluto – its diameter, the eccentricity of its orbit, its number of moons – and still disagree on the question “Is Pluto a planet”, one can agree on every characteristic of every Israel opponent and still disagree on the definitional question “Is opposing Israel anti-Semitic?”

(fact: it wasn’t until proofreading this essay that I realized I had originally written “Is Israel a planet?” and “Is opposing Pluto anti-Semitic?” I would like to see Jonathan Haidt hypnotize people until they can come up with positive arguments for those propositions.)

What’s the point of this useless squabble over definitions?

I think it’s about drawing a line between the concept “anti-Semitism” and “oppose Israel”. If your head is screwed on right, you assign anti-Semitism some very bad karma. So if we can stick a thick line between “anti-Semitism” and “oppose Israel”, then you’re going have very bad feelings about opposition to Israel and your General Factor Of Pro-Israeliness will go up.

Notice that this model *is transitive, but shouldn't be*.

That is, let's say we're arguing over the definition of anti-Semitism, and I say "anti-Semitism just means anything that hurts Jews". This is a dumb definition, but let's roll with it.

First, I load "anti-Semitism" with lots of negative affect. Hitler was anti-Semitic. The pogroms in Russia were anti-Semitic. The Spanish Inquisition was anti-Semitic. Okay, negative affect achieved.

Then I connect "wants to end the Israeli occupation of Palestine" to "anti-Semitism". Now wanting to end the Israeli occupation of Palestine has lots of negative affect attached to it.

It sounds dumb when you put it like that, but when you put it like "You're anti-Semitic for wanting to end the occupation" it's a pretty damaging argument.

This is *trying* to be transitive. It's trying to say "anti-occupation = anti-Semitism, anti-Semitism = evil, therefore anti-occupation = evil". If this were arithmetic, it would work. But there's no Transitive Property Of Concepts. If anything, concepts are more like sets. The logic is "anti-occupation is a member of the set anti-Semitic, the set anti-Semitic contains members that are evil, therefore anti-occupation is evil", which obviously doesn't check out.

(compare: "I am a member of the set 'humans', the set 'humans' contains the Pope, therefore I am the Pope".)

Anti-Semitism is generally considered evil because a lot of anti-Semitic things involve killing or dehumanizing Jews. Opposing the Israel occupation of Palestine doesn't kill or dehumanize Jews, so even if we call it "anti-Semitic" by definition, there's no reason for our usual bad karma around

anti-Semitism to transfer over. But by an unfortunate rhetorical trick, it does – you can gather up bad karma into "anti-Semitic" and then shoot it at the "occupation of Palestine" issue just by clever use of definitions.

This means that if you can come up with sufficiently clever definitions and convince your opponent to accept them, you can win any argument by default just by having a complex system of mirrors in place to reflect bad karma from genuinely evil things to the things you want to tar as evil. This is essentially the point I make in Words, Words, Words.

If we kinda tweak the definition of "anti-Semitism" to be "anything that inconveniences Jews", we can pull a trick where we leverage people's dislike of Hitler to make them support the Israeli occupation of Palestine – but in order to do that, we need to get everyone on board with our *slightly* non-standard definition. Likewise, the social justice movement insists on their own novel definitions of words like "racism" that don't match common usage, any dictionary, or etymological history – but which do perfectly describe a mirror that reflects bad karma toward opponents of social justice while making it impossible to reflect any bad karma back. Overreliance on this mechanism explains why so many social justice debates end up being about whether a particular mirror can be deployed to transfer bad karma in a specific case ("are trans people privileged?!") rather than any feature of the real world.

But they are hardly alone. Compare: "Is such an such an organization a *cult*?", "Is such and such a policy *socialist*?", "Is abortion or capital punishment or war *murder*?" All entirely about whether we're allowed to reflect bad karma from known sources of evil to other topics under discussion.

Look around you. Just look around you. Have you worked out what we're looking for? Correct. The answer is The Worst Argument In The World. Only now, we can explain why it works.

VIII.

From the self-esteem literature, I gather that the self is also a concept that can have good or bad karma. From the cognitive dissonance literature, I gather that the self is actively involved in maintaining good karma around itself through as many biases as it can manage to deploy.

I've mentioned this study before. Researchers make ~~victims~~ participants fill out a questionnaire about their romantic relationships. Then they pretend to "grade" the questionnaire, actually assigning scores at random. Half the participants are told their answers indicate they have the tendency to be very faithful to their partner. The other half are told they have very low faithfulness and their brains just aren't built for fidelity. Then they ask the ~~participants~~ victims their opinion on staying faithful in a relationship – very important, moderately important, or not so important?

There is a strong signal of people who are told they are bad at fidelity to state fidelity is unimportant, and another strong signal of people who are told they are especially faithful stating that fidelity is a great and noble virtue that must be protected.

The researchers conclude that people want to have high self-esteem. If I am terrible at fidelity, and fidelity is the most important virtue, that makes me a terrible person. If I am terrible at fidelity and fidelity doesn't matter, I'm fine. If I am great at fidelity, and fidelity is the most important virtue, I can feel pretty good about myself.

This doesn't seem too surprising. It's just the more subtle version of the effect where white people are a lot more likely to be white supremacists than members of any other race. Everyone likes to hear that they're great. The question is whether they can defend it and fit it in with their other ideas. The answer is "usually yes, because people are capable of pretty much any contortion of logic you can imagine and a lot that you can't".

I had a bad experience when I was younger where a bunch of feminists attacked and threatened me because of something I wrote. It left me kind of scarred. More importantly, the shape of that scar was a big anticorrelated line between self-esteem and the "feminism" concept. If feminism has lots of good karma, then I have lots of bad karma, because I am a person feminists hate. If feminists have lots of bad karma, then I look good by comparison, the same way it's pretty much a badge of honor to be disliked by Nazis. The result was a permanent haze of bad karma around "feminism" unconnected to any specific feminist idea, which I have to be constantly on the watch for if I want to be able to evaluate anything related to feminism fairly or rationally.

Good or bad karma, when applied to yourself, looks like high or low self-esteem; when applied to groups, it looks like high or low status. In the giant muddle of a war for status that we politely call "society", this makes beliefs into weapons and the karma loading of concepts into the difference between lionization and dehumanization.

The Trope Namer for emotivist arguments is "ethnic tension", and although it's most obvious in the case of literal ethnicities like the Israelis and the Palestinians, the ease with which concepts become attached to different groups creates a whole lot of "proxy ethnicites". I've written before about how

American liberals and conservatives are seeming less and less like people who happen to have different policy prescriptions, and more like two different tribes engaged in an ethnic conflict quickly approaching Middle East level hostility. More recently, a friend on Facebook described the-thing-whose-name-we-do-not-speak-lest-it-appear and-destroy-us-all, the one involving reproductively viable worker ants, as looking more like an ethnic conflict about who is oppressing whom than any real difference in opinions.

Once a concept has joined up with an ethnic group, either a real one or a makeshift one, it's impossible to oppose the concept without simultaneously lowering the status of the ethnic group, which is going to start at least a *little* bit of a war. Worse, once a concept has joined up with an ethnic group, one of the best ways to argue against the concept is to dehumanize the ethnic group it's working with. Dehumanizing an ethnic group has always been easy – just associate them with a disgust reaction, portray them as conventionally unattractive and unlovable and full of all the worst human traits – and now it is profitable as well, since it's one of the fastest ways to load bad karma into an idea you dislike.

IX.

According to The Virtues Of Rationality:

> The tenth virtue is precision. One comes and says: The quantity is between 1 and 100. Another says: the quantity is between 40 and 50. If the quantity is 42 they are both correct, but the second prediction was more useful and exposed itself to a stricter test. What is true of one apple may not be true of another apple; thus more can be said about a single apple than about all the apples in the world. The narrowest statements slice deepest, the cutting

> edge of the blade. As with the map, so too with the art of mapmaking: The Way is a precise Art. Do not walk to the truth, but dance. On each and every step of that dance your foot comes down in exactly the right spot. Each piece of evidence shifts your beliefs by exactly the right amount, neither more nor less. What is exactly the right amount? To calculate this you must study probability theory. Even if you cannot do the math, knowing that the math exists tells you that the dance step is precise and has no room in it for your whims.

The official desciption is of literal precision, as specific numerical precision in probability updates. But is there a secret interpretation of this virtue?

> Four top secret Virtues known only to the Highest Clergy: 1) Fnorg 2) Turlity 3) Charigrace 4) Love-231.
>
> — Deity Of Religion (@deityofreligion) October 24, 2014

Precision as separation. Once you're debating "religion", you've already lost. Precision as sticking to a precise question, like "Is the first chapter of Genesis literally true?" or "Does Buddhist meditation help treat anxiety disorders?" and trying to keep these issues as separate from any General Factor Of Religiousness as humanly possible. Precision such that "God the supernatural Creator exists" and "God the order and beauty in the Universe exists" are as carefully sequestered from one another as "Did the defendant kill his wife?" and "Did the defendant kill the President?"

I want to end by addressing a point a commenter made in my last post on motte-and-bailey:

> In the real world, the particular abstract questions aren't what matter – the groups and people are what matter. People get things done, and they aren't particularly married to particular abstract concepts, they are married to their values and their compatriots. In order to deal with reality, we must attack and defend groups and individuals. That does not mean forsaking logic. It requires dealing with obfuscating tactics like those you outline above, but that's not even a real downside, because if you flee into the narrow, particular questions all you're doing is covering your eyes to avoid perceiving the the monsters that will still make mincemeat of your attempts to change things.

I don't entirely disagree with this. But I think we've been over this territory before.

The world is a scary place, full of bad people who want to hurt you, and in the state of nature you're pretty much obligated to engage in whatever it takes to survive.

But instead of sticking with the state of nature, we have the ability to form communities built on mutual disarmament and mutual cooperation. Despite artificially limiting themselves, these communities become stronger than the less-scrupulous people outside them, because they can work together effectively and because they can boast a better quality of life that attracts their would-be enemies to join them. At least in the short term, these communities can resist races to the bottom and prevent the use of personally effective but negative-sum strategies.

One such community is the kind where members try to stick to rational discussion as much as possible. These communities are definitely better able to work together, because they have a

powerful method of resolving empirical disputes. They're definitely better quality of life, because you don't have to deal with constant insult wars and personal attacks. And the existence of such communities provides positive externalities to the outside world, since they are better able to resolve difficult issues and find truth.

But forming a rationalist community isn't just about having the *will* to discuss things well. It's also about having the *ability*. Overcoming bias is really hard, and so the members of such a community need to be constantly trying to advance the art and figure out how to improve their discussion tactics.

As such, it's acceptable to try to determine and discuss negative patterns of argument, even if those patterns of argument are useful and necessary weapons in a state of nature. If anything, understanding them makes them *easier* to use if you've got to use them, and makes them easier to recognize and counter from others, giving a slight advantage in battle if that's the kind of thing you like. But moving them from unconscious to conscious also gives you the crucial *choice* of when to deploy them and allows people to try to root out ethnic tension in particular communities.

Race and Justice: Much More Than You Wanted to Know

Previously reviewed: effects of marijuana legalization, health effects of wheat, effectiveness of SSRIs, effectiveness of Alcoholics Anonymous

Does the criminal justice system treat African-Americans fairly?

I always assumed it obviously didn't. Then a while ago I read this harshly polemical but research-filled article claiming to prove it did. Then I found a huge review paper on the subject, written by a Harvard professor of sociology, which concluded after analyzing sixty pages of exquisitely-researched studies that:

> Recognizing that research on criminal justice processing in the United States is complex and fraught with methodological problems, the weight of the evidence reviewed suggests the following. When restricted to index crimes, dozens of individual-level studies have shown that a simple direct influence of race on pretrial release, plea bargaining, conviction, sentence length, and the death penalty among adults is small to nonexistent once legally relevant variables (e.g. prior record) are controlled. For these crimes, racial differentials in sanctioning appear to match the large racial differences in criminal offending. Findings on the processing of adult index crimes therefore generally support the non-discrimination thesis.

Clearly this was more complicated than I thought. I decided to waste my precious free time reading seven zillion contradictory studies to figure out what was going on. Some people on Tumblr have demanded I report back, so here goes:

A. Encounter Rate

There are a lot of tiers to the criminal justice system, each of which will have to be analyzed individual. The first tier is – who does or doesn't get stopped by the police?

One common point of discussion is traffic stops, leading to the popular joke that you can be stopped for a "DWB" (driving while black). Engel and Calnon (2006) seem to have done the definitive review in this area. Based on a national survey of citizens' interactions with police, they find that 5% of whites and 11% of blacks have had their cars searched by police, with relatively similar results for other kinds of officer interactions. Therefore, blacks are about twice as likely to be searched as whites. Once you do a multiple regression controlling for other factors, like previous record, income, area stopped, et cetera, half of that difference goes away, leaving an unexplained relative risk of 1.5x.

These data admit to multiple possible interpretations. First, racist police officers could be unfairly targeting blacks. Second, blacks could be acting more suspiciously and police officers correctly picking up on this fact. Third, police officers could be racially profiling based on their past experience of more successful searches of black drivers.

One common method of disentangling these possibilities is search "success rate". That is, if searching whites usually turns up more real crimes than searching blacks, then innocent blacks are being searched disproportionately often and the police are not just correctly responding to indicators of suspiciousness or past experiences.

Engel and Calnon review sixteen studies investigating this question. If we limit claims of dissimilarity to studies where one race is at least five percentage points higher than the other, there are eight studies with racial parity, six studies with higher white hit rates, and two studies with higher black hit rates.

In other words, in 62% of studies, police are not searching blacks disproportionately to the amount of crimes committed or presumed "indicators of suspiciousness". In 38% of studies, they are. The

differences may reflect either methodological differences (some studies finding effects others missed) or jurisdictionial differences (some studies done in areas where the police were racially biased, others done in areas where they weren't)

The authors did their own analysis based on a national survey about citizens' contact with the police, and found that 16% of whites searched and 8% of minorities searched reported that police had discovered contraband, a statistically significant difference. This contradicts the studies above, most of which found no difference and the others of which found much smaller differences.

One possible explanation the authors bring up is that previous research has shown black drivers who have received traffic violations are less likely than whites who have received traffic violations to admit to having received them on anonymous research surveys. For example, among North Carolina drivers known to have received tickets, 75% of whites admitted it on a survey compared to 66% of blacks (Pfaff-Wright, Tomaskovic-Devey, 2000). Comparisons of several different surveys of drug use find that "nonreporting of drug use is twice as common among blacks and Hispanics as among whites" (Mensch and Kandel). Since much of the "contraband" these surveys were asking about was, in fact, drugs, this seems pretty relevant. Overall different studies find different black-white reporting gaps (from the very small one in the traffic ticket study to the very large one on the drug use surveys). Plausibly this is related to severity of offense. Also plausibly, it relates to differential levels of trust in the system and worry about being found out – for poor black people, the possibility of (probably white) researchers being stooges who are going to send their supposedly confidential surveys to the local police station and get them locked up might be much more salient.

There are of course many other forms of police stop. These tend to follow the same pattern as traffic stops – strong data that police more often stop black people, police making the claim that black people do more things that trigger their suspicion instinct (including live in

higher-crime neighborhoods), and difficulty figuring out whether this is true or false.

Sampson and Lauritsen review several studies on police stops of pedestrians. I'll be coming back to and citing sources from this Sampson and Lauritsen article many times during this discussion as it is one of the most rigorous and trustworthy analyses around – Sampson is Professor of Sociology at Harvard and winner of the Stockholm Prize in Criminology and his review is the most cited one on this topic I could find, so I assume he represents something like a mainstream position. After reviewing a few studies, most notably Smith (1986), they conclude these sorts of police stops demonstrate no direct effect of race – in any given neighborhood, black people and white people are treated equally – but that there is an indirect effect from neighborhood – that is, the police are nastier to everybody in black neighborhoods. Although they don't say so, the most logical explanation to me would be that black neighborhoods are poorer and therefore higher crime, and so the police are more watchful and/or paranoid.

Summary: There is good data that police stop blacks more often, both on the road and in neighborhoods. Studies conflict over whether the extra stops are justifiable; likely this varies by jurisdiction. Extra neighborhood stops are most likely neighborhood-related effects rather than race-related per se, but the neighborhood effects do disproportionately target black people.

B. Arrest Rates For Violent Crimes

Police records consistently show that black people are arrested at disproportionally high rates (compared to their presence in the population) for violent crimes. For example, blacks are arrested eight times more often for homicide and fourteen times more often for robbery. Even less flashy crimes show the same pattern: forgery, fraud, and embezzlement all hover around a relative risk of four.

(White people are arrested at disproportionally high rates for things like driving drunk, and Asians are arrested at disproportionally high

rates for things like illegal gambling, but these carry lower sentences and are less likely to lead to incarceration.)

Once again, there are two possible hypotheses here: either police are biased, or black people actually commit these crimes at higher rates than other groups.

The second hypothesis has been strongly supported by crime victimization surveys, which show that the percent of arrestees who are black matches very closely matches the percent of victims who say their assailant was black. This has been constant throughout across thirty years of crime victmization surveys.

While everybody is totally on board with attributing this to structural factors like black people being poorer and living in worse neighborhoods, anyone who tries to analyze higher black arrest and incarceration rates without taking this into account is going to end up extremely confused.

There were some attempts to cross-check police data and victim data against self-reports of criminality among different races, with various weird and wonderful results. Once again, after a while someone had the bright idea to check whether people who said they hadn't committed any crimes *actually* hadn't committed any crimes, and found that a lot of them had well-verified criminal records longer than *War And Peace*.

Sociologists learned an important lesson that day, which is that *criminals sometimes lie about being criminals*.

No one has had any better ideas for how to corroborate the crime victimization survey data, so it looks like probably that's the best we will do.

Summary: Arrests for violent crimes are probably not racially biased.

C. Arrest Rates For Minor Crimes

Usually when people talk about racial disparities in arrest rates for minor crimes, they're talking about drugs. The basic argument is that

black people and white people use drugs at “similar rates”, but black people are four times more likely to get arrested for drug crime. You can find this argument on pretty much every major media outlet: NYT, Slate, Vox, HuffPo, USA Today, et cetera.

The Bureau of Justice has done their own analysis of this issue and finds it’s more complicated. For example, all of these “equally likely to have used drugs” claims turn out to be that blacks and whites are equally likely to have “used drugs in the past year”, but blacks are far more likely to have used drugs in the past week – that is, more whites are only occasional users. That gives blacks many more opportunities to be caught by the cops. Likewise, whites are more likely to use low-penalty drugs like hallucinogens, and blacks are more likely to use high-penalty drugs like crack cocaine. Further, blacks are more likely to live in the cities, where there is a heavy police shadow, and whites in the suburbs or country, where there is a lower one.

When you do the math and control for all those things, you halve the size of the gap to “twice as likely”.

The Bureau of Justice and another source I found in the Washington Post aren’t too sure about the remaining half, either. For example, anecdotal evidence suggests white people typically do their drug deals in the dealer’s private home, and black people typically do them on street corners. My personal discussions with black and white drug users have turned up pretty much the same thing. One of those localities is much more likely to be watched by police than the other.

Finally, all of this is based on self-reported data about drug use. Remember from a couple paragraphs ago how studies showed that black people were twice as likely to fail to self-report their drug use? And you notice here that black people are twice as likely to be arrested for drug use as their self-reports suggest? That’s certainly an interesting coincidence.

The Bureau of Justice takes this possibility very seriously and adds:

> Although arrested whites and arrested blacks were about equally likely to be drug use deniers, these results nevertheless have implications for the SAMHSA survey. A larger fraction of the black population than the white population consists of criminally active persons and, therefore, a larger fraction of the black population than the white population would consist of criminally active persons who use drugs but deny it. Consequently, the SAMHSA survey would probably understate the difference between whites and blacks in terms of drug use. Whether the effect of such drug use denial among criminally active persons is large enough to account for the unexplained 13% is not known, but research on the topic should pursue this possibility.

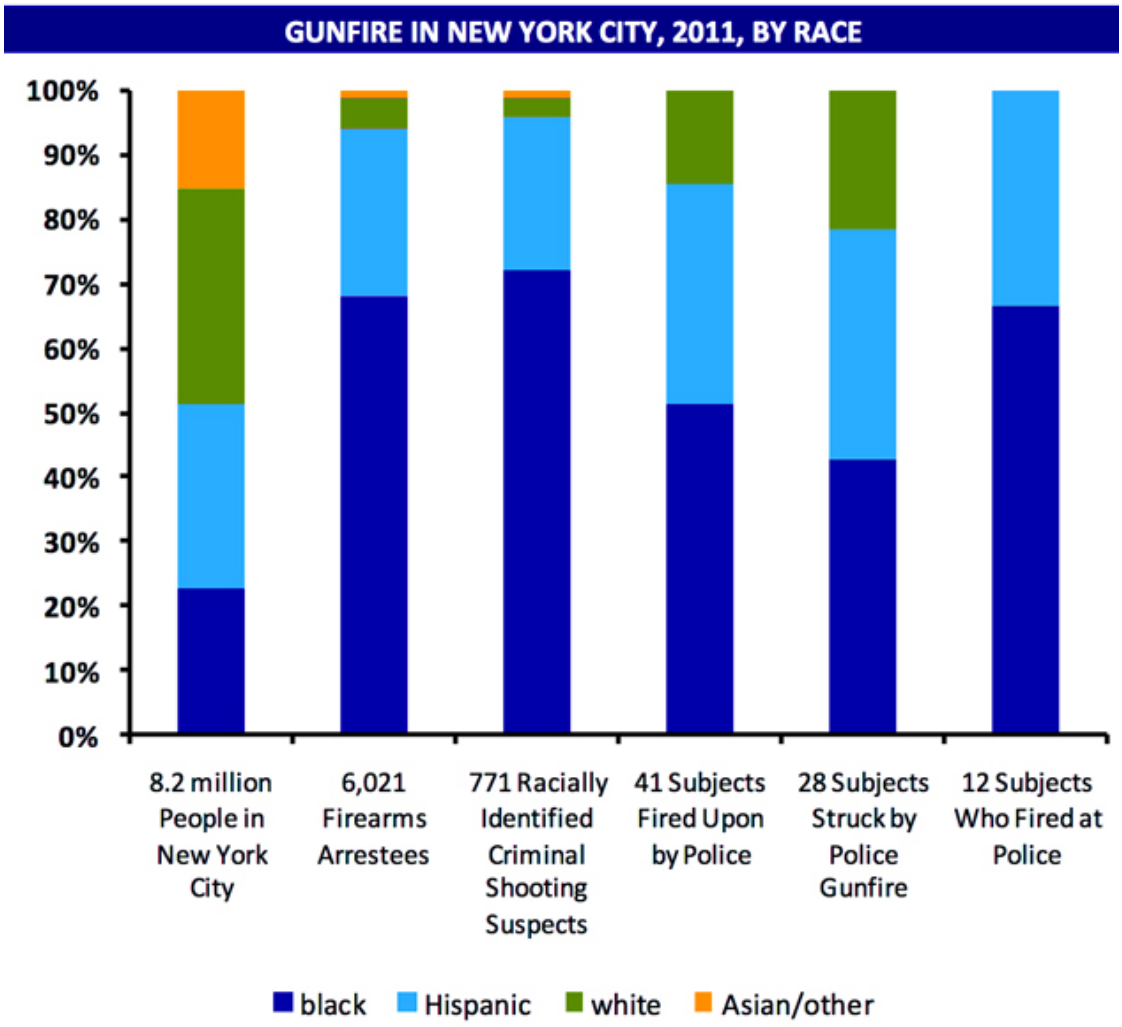
It should be noted that a study investigating this methodology gave random urine drug tests to some of the people who had filled out this survey, and found that half of the actual drug users had reported on the survey that they were squeaky clean. There were no racial data associated with this investigation, which is too bad.

Summary: Blacks appear to be arrested for drug use at a rate four times that of whites. Adjusting for known confounds reduces their rate to twice that of whites. However, other theorized confounders could mean that the real relative risk is anywhere between two and parity. Never trust the media to give you any number more complicated than today's date..

D. Police Shootings

A topical issue these days. Once again, the same dynamic at play. We know black people are affected disproportionately to their representation in the population, but is a result of police racism or disproportionate criminality?

Mother Jones magazine has an unexpectedly beautiful presentation of the data for us:



The fourth bar seems like what we're looking for. You could go with the fifth bar, but then you're just adding noise of who did or didn't duck out of the way fast enough.

As you can see, a person shot at by a police officer is more than twice as likely to be black as the average member of the general population. But, crucially, they are less likely to be black than the average violent shooter *or* the average person who shoots at the police.

We assume that the reason an officer shoots a suspect is because that officer believes the suspect is about to shoot or attack the officer. So if the officer were perfectly unbiased, then the racial distribution of people shot by officers would look exactly like the distribution of dangerous attackers. If it's blacker than the distribution of dangerous

attackers, the police are misidentifying blacks as dangerous attackers.

But In fact, the people shot by police are less black than the people shooting police or the violent shooters police are presumably worried about. This provides very strong evidence that, at least in New York, the police are not disproportionately shooting black people and appear to be making a special effort to avoid it.

For some reason most of the studies I could get here were pretty old, but with that caveat, this is also the conclusion of Milton (1977) looking at police departments in general, and Fyfe (1978), who analyzes older New York City data and comes to the same conclusion. However, the same researcher analyzes police shootings in Memphis and finds that these *do* show clear evidence of anti-minority bias, sometimes up to a 6x greater risk for blacks even after adjusting for likely confounders. The big difference seems to be that NYC officers are trained to fire only to protect their own lives from armed and dangerous suspects, but Memphis officers are (were? the study looks at data from 1970) allowed to shoot property crime suspects attempting to flee. The latter seems a lot more problematic and probably allows more room for officer bias to get through.

[EDIT: A commenter pointed out to me that *Tennessee vs. Garner* banned this practice in the late 1980s, meaning Memphis' shooting rate should be lower and possibly less biased now]

The same guy looks at the race of officers involved and finds that "the data do not clearly support the contention that white [officers] had little regard for the lives of minorities". In fact, most studies find white officers are disproportionately more likely to shoot white suspects, and black officers disproportionately more likely to shoot black suspects. This makes sense since officers are often assigned to race-congruent neighborhoods, but sure screws up the relevant narrative.

Summary: New York City data suggests no bias of officers towards shooting black suspects compared with their representation among dangerous police encounters, and if anything the reverse effect. Data

from Memphis in 1970 suggests a strong bias towards shooting black suspects, probably because they shoot fleeing suspects in addition to potentially dangerous suspects, but this practice has since stopped. Older national data skews more toward the New York City side with little evidence of racial bias, but I don't know of any recent studies which have compared the race of shooting victims to the race of dangerous attackers on a national level. There is no support for the contention that white officers are more likely than officers of other races to shoot black suspects.

E. Prosecution And Conviction Rates

Conviction rates of blacks have generally found to be less than than conviction rates of whites (Burke and Turk 1975, Petersilia 1983, Wilbanks 1987). I don't know why so many of these studies are from the 70s and 80s, but a more recent Bureau of Justice Statistics finds that 66% of accused blacks get prosecuted compared to 69% of accused whites; 75% of prosecuted blacks get convicted compared to 78% of prosecuted whites.

The 1975 study suggested this was confounded by type of crime – for example, maybe blacks are charged more often with serious crimes for which the burden of proof is higher. The 1993 study isn't so sure; it breaks crimes down by category and finds that if anything the pro-black bias becomes *stronger*. For example, 51% of blacks charged with rape are acquitted, compared to only 25% of whites. 24% of blacks charged with drug dealing are acquitted, compared to only 14% of whites. Of fourteen major crime categories, blacks have higher acquittal rates in twelve of them (whites win only in "felony traffic offenses" and "other").

The optimistic interpretation is that there definitely isn't any sign of bias against black people here. The pessimistic interpretation is that this would be consistent with more frivolous cases involving black people coming to the courts (ie police arrest blacks at the drop of a hat, and prosecutors and juries end up with a bunch of stupid cases without any evidence that they throw out).

There was a much talked-about study recently that found that "juries were equally likely to convict black and white offenders when there was at least one black in the jury pool, but more likely to convict blacks when there wasn't." This is consistent with previous studies. Jury pools contain twenty-seven members; the probability that there will be at least one black jury pool member in the trial of a black subject (who of course is most likely to live in a predominantly black area) is high. The study's "equally likely to convict black and white offenders" was actually "2% more likely to convict white offenders than black offenders", which was probably not statistically significant with its small sample size but is consistent with the small pro-black effects found elsewhere.

Summary: Prosecution and conviction rates favor blacks over whites, significance unclear.

F. Sentencing

Older studies of sentencing tend to find no or almost no discrepancies between blacks and whites. This was the conclusion of most of the papers reviewed in Sampson and Lauritsen. The gist here seems to be that there were "four waves" of studies in this area. The first wave, in the 1960s, was naive and poorly controlled and found that there was a lot of racial bias. The second wave, in the 1980s, controlled for more things (especially prior convictions) and found there wasn't. The third wave was really complicated, and the writers sum it up as saying it represented:

> …a shift away from the non-discrimination thesis to the idea that there is *some* discrimination, *some* of the time, in *some* places. These contingencies undermine the broad reach of the thesis, but the damage is not fatal to the basic argument that race discrimination is not pervasive or systemic in criminal justice processing.

The fourth wave expands on this and finds discrimination in some areas that hadn't been studied before, such as plea bargaining. However, it continues to find that on the whole, and especially in the

largest and best-designed studies there is very little evidence of discrimination. The article concludes:

> Langan's interpretation matches those of other scholars such as Petersilia (1985) and Wilbanks (1987) in suggesting that systemic discrimination does not exist. Zatz (1987) is more sympathetic to the thesis of discrimination in the form of indirect effects and subtle racism. But the proponents of this line of reasoning face a considerable burden. If the effects of race are so contingent, interactive, and indirect in a way that to date has not proved replicable, how can one allege that the "system" is discriminatory?

A more recent (fifth wave?) review adds some problems to this generally rosy picture, saying that "Of the [thirty-two studies containing ninety-five different] estimates of the direct effect of race on sentencing at the state level, 43.2% indicated harsher sentences for blacks…at the federal level 68.2% of the [eight studies containing twenty-two different] estimates of the direct effect of race on sentencing indicated harsher sentences for blacks". The majority of estimates that did not find this were race-neutral, although six did show some bias against whites. They conclude:

> Racial discrimination in sentencing in the United States today is neither invariable nor universal, nor is it as overt as it was even thirty years ago. As will be described below, while the situation has improved in some ways, racially discriminatory sentencing today is far more insidious than in the past, and treating a racial or ethnic group as a unitary body can mask the presence of discrimination.

I really like how you can make a large decrease in the level of a bad thing sound like a problem by saying it is becoming "more insidious".

Even more recent studies have found even larger gaps. A study by the US Sentencing Commission investigating the effect of new

guidelines found that blacks' sentences were 20% longer than those of similar whites; a later methodological update reduced the gap to a still-large 14.5% and a a different recent study says just under 10%. Although the particular effect of these new guidelines is a matter of HORRIBLE SUPER-COMPLICATED DEBATE, neither side seems to deny the disparities themselves – only whether they are getting larger.

It's not clear to me why there's such a difference between the earlier studies (which found little evidence of disparity), the middle studies (which were about half-and-half), and these later studies (which show strong evidence of disparity). I guess one side of a HORRIBLE SUPER-COMPLICATED DEBATE would say it has to do with changes in sentencing during that time which replace mandatory sentences with "judicial discretion". If you're mandated to give a particular sentence for a particular crime, there's a lot less opportunity to let bias slip in then if you can do whatever you want. There is some evidence that different judges treat different races differently, although the study has no way of proving whether this is anti-black bias, anti-white bias, or an equal mix of both in different people. Unfortunately, there is also concern that mandatory minimum sentencing is itself racist.

Capital punishment is in its own category, and pretty much all studies, old, new, anything agree it is racist as heck (Sampson and Lauritsen cite Bowers & Pierce 1980; Radelet 1981; Paternoster 1984; Keil and Vito 1989; Aguirre and Baker 1990; Baldus Woodward & Pulaski 1990 – there's no way I'm reading through all of them so I will trust they say what the review says they say). This seems to consist not only in black suspects being more at risk, but in white victims' deaths being more likely to get their offenders a death sentence.

Summary: Most recent studies suggest a racial sentencing disparity of about 15%, contradicting previous studies that showed lower or no disparity. Changes in sentencing guidelines are one possible

explanation; poorly understood methodological differences are a second. Capital punishment still sucks.

Summary

There seems to be a strong racial bias in capital punishment and a moderate racial bias in sentence length and decision to jail.

There is ambiguity over the level of racial bias, depending on whose studies you want to believe and how strictly you define "racial bias", in police stops, police shootings in certain jurisdictions, and arrests for minor drug offenses.

There seems to be little or no racial bias in arrests for serious violent crime, police shootings in most jurisdictions, prosecutions, or convictions.

Overall I disagree with the City Journal claim that there is no evidence of racial bias in the justice system.

But I also disagree with the people who say things like "Every part of America's criminal justice is systemically racist by design" or "White people can get away with murder but black people are constantly persecuted for any minor infraction," or "Every black person has to live in fear of the police all the time in a way no white person can possibly understand". The actual level of bias is limited and detectable only through statistical aggregation of hundreds or thousands of cases, is only unambiguously present in sentencing, and there only at a level of 10-20%, and that only if you believe the most damning studies.

(except that you should probably stay out of Memphis)

It would be nice to say that this shows the criminal justice system is not disproportionately harming blacks, but unfortunately it doesn't come anywhere close to showing anything of the sort. There are still many ways it can indirectly harm blacks without being explicitly racist. Anatole France famously said that "the law, in its majestic equality, forbids rich *as well as* poor people from begging for bread and sleeping under bridges", and in the same way that the laws France cites, be they enforced ever so fairly, would still

disproportionately target poor people, so other laws can, even when fairly enforced, target black people. The classic example of this is crack cocaine – a predominantly black drug – carrying a higher sentence than other whiter drugs. Even if the police are scrupulously fair in giving the same sentence to black and white cokeheads, the law will still have a disproportionate effect.

There are also entire classes of laws that are much easier on rich people than poor people – for example, any you can get out of by having a good lawyer – and entire classes of police work that are harsher on poor neighborhoods than rich neighborhoods. If the average black is poorer than the average white, then these laws would have disproportionate racial effects.

For more information on this, I would recommend Tonry and Melewski's *Malign Neglect: Race, Crime, and Punishment in America*. They begin by saying everything above is true – the system mostly avoids direct racist bias against black people – and go on to say argue quite consistently that we *still* have a system where (their words) "recent punishment policies have replaced the urban ghetto, Jim Crow laws, and slavery as a mechanism for maintaining white dominance over blacks in the United States". If you want something that makes the strongest case for the justice system harming blacks, written by real criminologists who know what they're talking about, there's your best bet.

(warning: I haven't read the book. I did read a review article by the same people, which the book is partially based on)

Some police officers say the reason they are harsher in poor urban neighborhoods is that the expectation of high levels of unruly behavior necessitates unusually strong countermeasures. For the same reason, I am screening all comments for the next few days. If you post one, expect it to show up eventually or perhaps disappear into the aether.

Framing for Light Instead of Heat

I.

Ezra Klein uses my analysis of race and justice as a starting point to offer a thoughtful and intelligent discussion of what exactly it means to control for something in a study.

I'm not really going to call it a critique of my piece, because it only applies to two of the six areas I looked at, and in those two areas Klein's thoughts were already carefully integrated into my conclusion – I described both as showing "ambiguity over the level of racial bias, depending on…how strictly you define racial bias." The Vox article repeated and expanded on that conclusion rather than contradicting it.

But it's still an important issue and I'm glad it's come up since I didn't have time to deal with at enough length in the original post.

The argument is: any study worth its salt is going to control for things like income level. Therefore, a study that concludes "blacks and whites get arrested at about the same rates" may only mean "blacks and whites of the same income level get arrested at about the same rates". If blacks on average have lower incomes, then in the real world blacks might still be arrested much more. Blacks being poor and therefore getting uniquely poor treatment from the criminal justice system (Klein says) sounds like *exactly* the sort of thing we would call "racial discrimination" or "racial bias" or "racism", but it would be totally missed by the standard methodology of controlling for income.

The solution is terminological rigor, which I foolishly forgot to have. What I should have said at the beginning of my post was

“I want to know whether there is any direct bias against black people caused by racist attitudes among police and other officials.” By this definition, all of my conclusions stand.

Klein wants to know whether there is any factor at all that causes disproportionate impact of the criminal justice system on any race. By this definition, my conclusions are only a tiny part of the picture, although at the end I recommend the book Malign Neglect which provides much of the rest.

As long as we keep these two meanings of “racial bias” or “racism” or whatever separate, there’s no problem. Once we start conflating them, we’re going to become very confused in one direction or another.

Ezra Klein and I don’t disagree about any point of statistics. What I think we do disagree about is the terminology.

If we find that much of the overrepresentation of blacks in the criminal justice system is because black people are often poor and poor people often get sucked into the system, should we describe this as “the problem isn’t racism in the criminal justice system, it’s poverty” or as “the problem is racism in the criminal justice system, as manifested through poverty”?

II.

Consider a town with 1000 black people and 1000 white people. 750 black people are poor, and 250 are rich. 750 white people are rich, and 250 are poor. Everyone commits crimes at the same rate – let’s say 10% per year. Rich people have lots of connections and can bribe their way out of trouble in a pinch, so only 50% of rich criminals get arrested. Poor people don’t have any strings they can pull, so 100% of poor criminals get arrested.

We can do the calculations and determine that the black arrest rate will be 8.75% and the white arrest rate 6.25%, a pretty significant difference. The people in the town can do the calculations as well. They correctly observe that in their town, everyone commits crimes at the same rate, so there must be some bias in their system. Using Klein's definition, they determine that since the system in their town disproportionately affects blacks, their criminal justice system is racist.

The problem is, upon learning that your criminal justice system is racist, what solutions come to mind? The ones I think of include things like increasing the diversity of the officer pool, sending police to diversity training, ferreting out racist attitudes and comments among members of the force, urging officers to consume media that is more positive towards black people, et cetera.

But all of these are unrelated to the problem and will accomplish nothing. We specified the decision algorithm these officers use, and we know it has nothing to do with race and everything to do with class. The townspeople should be attacking the culture of bribery, nepotism, and corruption, not throwing away resources on curing racist attitudes that don't affect police behavior in the slightest.

Note that this is true *even if* the poverty is caused by racism. Suppose the town college unfairly admits whites and turns down blacks, which is why the white people in this town are so much richer. I have no problem with saying "the town college is racist". This suggests the appropriate solutions – educating and/or punishing the people at the college. I have a lot of problems with saying "the town police are racist" as a shortcut for "the town police take bribes, and due to racism

somewhere else the people with the cash are all white" because this obfuscates the correct solution.

You can't just cut links out of a causal chain and preserve meaning. "Blacks are arrested disproportionately often because of gravity" is true, insofar as without the formation of the Earth from the gravitational coalescence of a primordial gas cloud humans and therefore racism wouldn't exist. But if the natural reaction to hearing the phrase is to solve the problem by attaching hundreds of helium balloons to black people, then say something less misleading.

III.

Klein goes on to say:

> An example is research around the gender wage gap, which tries to control for so many things that it ends up controlling for the thing it's trying to measure. As my colleague Matt Yglesias wrote, the commonly cited statistic that American women suffer from a 23 percent wage gap through which they make just 77 cents for every dollar a man earns is much too simplistic. On the other hand, the frequently heard conservative counterargument that we should subject this raw wage gap to a massive list of statistical controls until it nearly vanishes is an enormous oversimplification in the opposite direction. After all, for many purposes gender is itself a standard demographic control to add to studies — and when you control for gender the wage gap disappears entirely!
>
> The question to ask about the various statistical controls that can be applied to shrink the gender gap is what are they actually telling us," he continued. "The answer, I think, is that it's telling how the wage gap works.

> Take hours worked, which is a standard control in some of the more sophisticated wage gap studies. Women tend to work fewer hours than men. If you control for hours worked, then some of the gender wage gap vanishes. As Yglesias wrote, it's "silly to act like this is just some crazy coincidence. Women work shorter hours because as a society we hold women to a higher standard of housekeeping, and because they tend to be assigned the bulk of childcare responsibilities."
>
> Controlling for hours worked, in other words, is at least partly controlling for how gender works in our society. It's controlling for the thing that you're trying to isolate.

Once again, when someone says "women make seventy seven cents for each dollar a man earns", the response is almost always "That's outrageous!" and demands that companies stop being so sexist. I don't even have to speculate here. Google "gender wage gap", and just on the first page of results you find statements like:

"While some CEOs have been vocal in their commitment to paying workers fairly, American women can't wait for trickle-down change. The American Association of University Women urges companies to conduct salary audits to proactively monitor and address gender-based pay differences."

"Our project on sex and race discrimination in the workplace shows that outright discrimination in pay, hiring, or promotions continues to be a significant feature of working life…the Institute for Women's Policy Research examined organizational remedies such as sexual harassment training, the introduction of new grievance procedures, supervisory training or revised performance management, and reward schemes."

"Today marks Equal Pay Day, the date that symbolizes how far into the new year the average American woman would have to work to earn what the average American man did in the previous year. With a new executive order issued today, President Obama and Democrats are hoping to peg the gender wage gap as a major issue ahead of the 2014 elections. This week, Senate Democrats also plan to again bring forward the proposed "Paycheck Fairness Act," a bill that aims to eliminate the pay gap between female and male employees. Both men and women see a need for moves such as this – 72% of women and 61% of men said "this country needs to continue making changes to give men and women equality in the workplace"

Given that the supposed gender pay gap is being used at this very moment to argue for salary audits, sexual harassment training, grievance procedures, and paycheck fairness acts, isn't it really important to know that a lot of it is due to upstream factors like how men and women are socialized as children to have different values, which wouldn't be affected by these things at all?

(Given that the entire issue is probably being used to load the term "feminism" with positive affect, isn't it important to know that it's mostly unrelated to what we expect feminists to do with their extra trust and power?)

It might be worthwhile to come at this from an ideologically opposite angle. Suppose I state "Professors who identify as feminist give twice as many As to female students as they do to male students."

(This is true, by the way.)

It sounds like a big problem. So you dig through mountains of data, and you figure out that most feminist professors tend to

be in subjects like the humanities, where twice as many students are female as male, and so naturally twice as many of the As go to women as men. If I just give you my best trollface and say "Yes, that's certainly the *mechanism* by which the extra female As occur", you have every reason to believe I'm deliberately causing trouble. Especially if colleges have already vowed to stop hiring feminist professors in response to the subsequent outrage. Especially especially if you know I am a cultural conservative activist whose goal has always been to make colleges stop hiring feminist professors, by hook or by crook.

If twice as many women as men take English literature classes, that's compatible with something about gender socialization unfairly making men feel less able to study or less enthusiastic about studying literature. That could be considered biased or discriminatory, I guess. But phrasing it as "feminist professors give twice as many As to women" is calculated to produce maximal damage. It's the sort of thing you would only do if you wanted to throw a match on a gunpowder keg for s**ts and giggles.

IV.

So I guess I've moved on from "poor choice of terminology" to "active misrepresentation". Let's stick with that.

This issue makes for the ultimate motte-and-bailey doctrine.

You go around saying "Gender gap! Women making less than men! Discrimination! Sexism!" Everyone puts on their Gricean implicature caps and concludes that they mean what these words mean in everyday speech. The appropriate remedies are trotted out – companies need to raise their female employees' pay, companies need to hire more discrimination officers, feminists need to talk more about all the ways men

talk over women in the workplace and mansplain to them, etc. This is the bailey.

Then someone says “Wait, according to our study, a lot of this is just that women prefer working shorter hours to have time with their families” – and so they retreat to their motte: “Yeah, that’s the *mechanism* for the gender gap. You mad, bro?”

But the thing about mottes is that nobody actually cares about them when there’s this awesome bailey they can fight over instead. By turning differential socialization into the motte for sexual harassment or something, we’re doing a disservice not only to sexual harassment, but to the principled study of differential socialization.

Anyway, the situation is actually even worse than this. If you hear “The problem with the criminal justice system is disproportionate impact on the poor,” then you’ll probably start coming up with ideas for how to deal with that, and other people will probably start listening. If you hear “the problem with the criminal justice system is racism,” then you will start sharpening your knives.

Racism is a *uniquely* divisive issue. Minorities hear it and think of Klansmen trying to kill them. White people hear it and think of witch-hunters trying to get them fired. A single death in a random Midwestern town has turned half the country into experts on ballistics because it involved race. Bring up race, and people will change their opinion in the opposite direction suggested by the evidence just to spite you for having a different opinion about it than they do.

Once you’ve said words like “racism” or “racial bias”, this dynamic is already in play and you have lost control of the conversation from then on. If you mention the word and then suggest that we should do something about the police bribery

or whatever, then ten percent are going to yell “HOW DARE YOU IMPUGN OUR OFFICERS’ HONOR, YOU POLITICALLY CORRECT FASCIST”, another ten percent are going to yell “HOW DARE YOU DERAIL THE CONVERSATION ABOUT RACE, YOU WHITE SUPREMACIST ASSHOLE”, and the other eighty percent are going to be yelling so loud at each other they can’t even hear you. By the time all the fires have been put out and all the rubble cleared, it’s a pretty good bet that nobody is in the mood to hear about policy ideas for reducing the impact of police on lower-income individuals anymore.

Klein ends his piece by interviewing a professor who states that “Liberals sometimes overstate the extent of overt racism as a direct explanation of justice system disparities.” He acts like this is some sort of inexplicable quirk of the liberal mind. I wonder whether it might have more to do with liberals reading things like the recent Vox article, “America’s Criminal Justice System Is Racist”, which declares the thesis “There is no reason to be subtle on this point: the American criminal justice system is racist”, then goes on to repeat the phrase “America’s criminal justice system is racist” five times in the next five paragraphs. It never mentions that possibility that any of this racism is anything but overt.

If, like Robin Hanson, you believe in the metaphor of tugging policy ropes sideways, then I can’t think of any worse way to ensure that everyone will be tugging against you in every direction than trying to focus the discussion about race.

That’s why I limited my review to direct bias within the justice system itself, and why I think other ways of framing the issue are less productive.

(Comment screening is on again, I guess. Comments that will start flame wars or derail the conversation will vanish into the aether. Unrelated: the book review yesterday got popular and this blog might go down every so often because of too much traffic. It’ll be up again shortly.)

The Wonderful Thing About Triggers

[Content note: hypothetical spiders]

I complain a lot about the social justice movement. Or for a change, I sometimes complain that the media is too friendly to the social justice movement. So when the media starts challenging the movement, with articles like Trigger Warnings: New Wave Of Political Correctness and We've Gone Too Far With Trigger Warnings and Warning: The Literary Canon Could Make Children Squirm and America's College Kids Are A Bunch Of Mollycoddled Babies, I really ought to be happy things are finally going my way.

Instead I'm a little disturbed. Let's fnord that last article:

> poor dears demand riot in the streets shield their precious eyes anything potentially offensive cave in, the most sacrosanct doctrines are endangered, buildings being "occupied," professors intimidated, deans confronted, generalized kindling of political correctness, self-absorption, spoiled-bratism, kids accustomed to getting their own way with just about everything, hovered over and indulged by their parents, grade-inflated carefully cushioned, precious as they are, schizy and spoiled, crop of prissy, protected and self-absorbed young people, shelter them from everything they don't already believe and welcome

This doesn't look good. Also, Jezebel and Baffler are against trigger warnings, as are a group of professors who teach "gender, sexuality, and critical race studies" (the last of which deals twice as much damage as regular race studies). Reversed

stupidity is not intelligence, but sometimes it's a helpful clue about where to look.

I like trigger warnings. I like them because they're not censorship, they're the opposite of censorship. Censorship says "Read what we tell you". The opposite of censorship is "Read whatever you want". The philosophy of censorship is "We know what is best for you to read". The philosophy opposite censorship is "You are an adult and can make your own decisions about what to read".

And part of letting people make their own decisions is giving them relevant information and trusting them to know what to do with them. Uninformed choices are worse choices. Trigger warnings are an attempt to provide you with the information to make good free choices of reading material.

And my role model here, as in so many other places, is Commissioner Lal: "Beware he who would deny you access to information, for in his heart, he dreams himself your master."

Trigger warnings fight those who would like to be our masters in another way as well. They are one of our strongest weapons against the proponents of censorship. The proponents say "We can't let you air that opinion, it might offend people." Trigger warnings say "I am explaining to you exactly how this might offend you, so if you continuing listening to me you have volunteered to hear whatever I have to say, on your own head be it, and let no one else purport to protect you from yourself."

I agree that bad people could use trigger warnings to avoid ever reading anything that challenges their prejudices. This is a problem with providing people informed choices. Sometimes they misuse them.

But I could also imagine good people using trigger warnings to *increase* their ability to read things that challenge their

views. Suppose you are a transgender person who becomes really uncomfortable when you hear people insult transgender people. Gradually you learn that a lot of people outside the social justice community do this a lot, so you stop reading anything outside the social justice community, forget about genuinely rightist sources like National Review or American Conservative. Now suppose sources start trigger warning their content. Most right-wing arguments *don't* insult transgender people, so all of a sudden you have a way to steer clear of the ones that do and read all of the others free from fear.

Actually, "fear" is the wrong word, it buys into the stereotyping of triggered people as coddled or cowards or something. Maybe some people feel fear. Others would just be free from exasperation, anger, distracting dismay, the cognitive load of having to hear people insult you and not being able to respond and having to exert effort to continue to read. I feel like this might be my response to the existence of more trigger warnings (at least if anyone ever warned for my triggers, which they won't).

And I guess I admit that the people who use trigger warnings for epistemic evil will probably outnumber those who use them for epistemic virtue. But then the question is: do "we", as a civilization, grant ourselves the right to force people to be virtuous without their consent? There are a lot of good arguments that we should, but that doesn't matter, because it's not a going question. In every other area of life, we've already decided that we don't. Like, it would be a spectacularly good idea to make a rule that every fifth link to Paul Krugman's blog has to redirect people to Tyler Cowen's blog, and vice versa, so people don't get a chance to only read the opinions they agree with. Or that every Republican has to watch one Daily Show a month, and every Democrat has to listen to one

Fox News segment. But if we're not going to do that, it hardly seems fair to put the whole burden of epistemic virtue on the easily triggered.

II.

The strongest argument against trigger warnings that I have heard is that they allow us to politicize ever more things. Colleges run by people on the left can slap big yellow stickers on books that promote conservative ideas, saying "THIS BOOK IS RACIST AND CLASSIST", and then act outraged if anyone requests a trigger warning that sounds conservative – like a veteran who wants one on books that vilify or mock soldiers, or a religious person who wants one on blasphemy. Then everyone has to have a big fight, the fight makes everyone worse off than *either* possible resolution, and it ends with somebody feeling persecuted and upset. In other words, it's an intellectual gang sign saying "Look! We can demonstrate our mastery of this area by only allowing our symbols; your kind are second-class citizens!"

On the other hand, this is terribly easy to fix. Put trigger warnings on books, but put them on the bullshytte page. You know, the one near the front where they have the ISBN number and the city where the publishers' head office is and something about the Library of Congress you've never read through even though it's been in literally every book you've ever seen. Put it there, on a small non-colorful sticker. Call it a "content note" or something, so no one gets the satisfaction of hearing their pet word "trigger warning". Put a generally agreed list of things – no sense letting every single college have its own acrimonious debate about it. The few people who actually get easily triggered will with some exertion avoid the universal human urge to flip past the bullshytte page and

spend a few seconds checking if their trigger is in there. No one else will even notice.

Or if it's about a syllabus, put it on the last page of the syllabus, in size 8 font, after the list of recommended reading for the class. As a former student and former teacher, I *know* no one reads the syllabus. You have to be really devoted to avoiding your trigger. Which is exactly the sort of person who should be able to have a trigger warning while everyone else goes ahead with their lives in a non-political way.

I'm sure there are some more implementation details, but it's nothing a little bit of good faith can't take care of. If good faith is used and some people *still* object because it's not EXACTLY what they want, *then* I'll tell them to go fly a kite, but not before.

I know a lot of people worry about slippery slopes; give the culture warriors an inch and they'll take a mile. I think this is a very backwards way of looking at things. Like, the anti-gay people talked about a slippery slope and fought desperately hard against gay marriage, even though it was pretty hard to find anything actually objectionable about it other than that it might be on a slippery slope to worse things. That desperate fight didn't delay gay marriage more than a few years, and it didn't prevent whatever gay marriage was on a slippery slope to. What it did do was *totally discredit conservatives in this area*. Now any time anyone makes a family values argument, even a good family values argument, people can say that "family values" is code for homophobia, and bring up that family values conservatives really *have* held abhorrent positions in the past so why should we trust them now? It gave liberals huge momentum, and if there is a slippery slope then all that opposing gay marriage did was destroy the credibility of anybody who could have stopped us going down it.

Opposing a good idea on slippery slope grounds is a moral failure *and* a strategic failure, and I'd hate for opponents of the social justice movement to make that mistake with trigger warnings.

III.

But this is all tangential to what really bothered me, which is Pacific Standard's The Problems With Trigger Warnings According To The Research.

You know, I love science as much as anyone, maybe more, but I have grown to dread the phrase "…according to the research".

They say that "Confronting triggers, not avoiding them, is the best way to overcome PTSD". They point out that "exposure therapy" is the best treatment for trauma survivors, including rape victims. And that this involves reliving the trauma and exposing yourself to traumatic stimuli, exactly what trigger warnings are intended to prevent. All this is true. But I feel like they are missing a very important point.

YOU DO NOT GIVE PSYCHOTHERAPY TO PEOPLE WITHOUT THEIR CONSENT.

Psychotherapists treat arachnophobia with exposure therapy, too. They expose people first to cute, little spiders behind a glass cage. Then bigger spiders. Then they take them out of the cage. Finally, in a carefully controlled environment with their very supportive therapist standing by, they make people experience their worst fear, like having a big tarantula crawl all over them. It usually works pretty well.

Finding an arachnophobic person, and throwing a bucket full of tarantulas at them while shouting "I'M HELPING! I'M HELPING!" works less well.

And this seems to be the arachnophobe's equivalent of the PTSD "advice" in the Pacific Standard. There are two problems with its approach. The first is that it avoids the carefully controlled, anxiety-minimizing setup of psychotherapy.

The second is that YOU DO NOT GIVE PSYCHOTHERAPY TO PEOPLE WITHOUT THEIR CONSENT.

If a person with post-traumatic stress disorder or some other trigger-related problem doesn't want psychotherapy, then *even as a trained psychiatrist* I am forbidden to override that decision unless they become an immediate danger to themselves or others.

And if they *do* want psychotherapy, then very likely they want to do it on their own terms. I try to read things that challenge my biases and may even insult or trigger me, but I do it *when I feel like it* and not a moment before. When I am feeling adventurous and want to become stronger in some way, I will set myself some strenuous self-improvement task, whether it be going on a long run or reading material I know will be unpleasant. But at the end of a really long and exasperating day when I'm at my wit's end and just want to relax, I don't want you chasing me with a sword and making me run for my life, and I don't want you forcing traumatic material at me.

The angry article above with all the talk of "spoiled brats" annoys me as an amateur politics blogger, but this Pacific Standard article pushes my buttons as a (somewhat) non-amateur psychiatrist. *This is not your job to meddle*. If you are very concerned about helping people with PTSD, please express that concern by donating to PTSD USA or one of the other organizations that will help those with the condition get proper, well-controlled therapy. Please do *not* try to increase

the background level of triggers in the hopes that one of them will fortuitously collide with a PTSD sufferer in a therapeutic way.

If, like me, you think the social justice movement has a really serious kindness and respect problem, then you know that it's really hard to bring this up without getting accused of unkindness and disrespect yourself. I don't know how to best respond to this problem. But I'm pretty sure that the very minimum one can do is not to *actually* be unkind and disrespectful. And I worry that some of these arguments against trigger warnings are failing to clear even this very low bar.

Fearful Symmetry

[Content warning: Social justice, anti-social justice, comparisons of social justice to anti-social-justice, comparisons of different groups' experiences.]

The social justice narrative describes a political-economic elite dominated by white males persecuting anybody who doesn't fit into their culture, like blacks, women, and gays. The anti-social-justice narrative describes an intellectual-cultural elite dominated by social justice activists persecuting anybody who doesn't fit into *their* culture, like men, theists, and conservatives. Both are relatively plausible; Congress and millionaires are 80% – 90% white; journalists and the Ivy League are 80% – 90% leftist.

The narratives share a surprising number of other similarities. Both, for example, identify their enemy with the spirit of a discredited mid-twentieth century genocidal philosophy of government; fascists on the one side, communists on the other. Both believe they're fighting a war for their very right to exist, despite the lack of any plausible path to reinstituting slavery or transitioning to a Stalinist dictatorship. Both operate through explosions of outrage at salient media examples of their out-group persecuting their in-group.

They have even converged on the same excuse for what their enemies call "politicizing" previously neutral territory – that what their enemies call "politicizing" is actually trying to restore balance to a field the other side has already successfully politicized. For example, on Vox recently a professor accused of replacing education with social justice propaganda in her classroom counterargues that:

> All of my students, regardless of the identity categories they embraced, had been taught their entire lives that real literature is written by white people. Naturally, they felt

> they were being cheated by this strange professor's "agenda"…It is worth asking, Who can most afford to teach in ways that are least likely to inspire controversy? Those who are not immediately hurt by dominant ideas. And what's the most dominant idea of them all? That the white, male, heterosexual perspective is neutral, but all other perspectives are biased and must be treated with skepticism […]
>
> Have we actually believed the lie that the only people who engage in "identity politics" are black feminists like me? Could it be that when some white men looked at more powerful white men, they could see them only as reasonable and not politically motivated, so they turned off their critical thinking skills when observing their actions? (Not everyone, of course.) Could it be that we only consider people ideologues when they don't vow allegiance to capitalism?

Compare to the "Sad Puppies", a group of conservatives accused of adding a conservative bent to science fiction's Hugo Awards. They retort that "politicization is what leftists call it when you fight back against leftists politicizing something". As per the Breitbart article:

> The chief complaint from the Sad Puppies campaigners is the atmosphere of political intolerance and cliquishness that prevails in the sci-fi community. According to the libertarian sci-fi author Sarah A. Hoyt, whispering campaigns by insiders have been responsible for the de facto blacklisting of politically nonconformist writers across the sci-fi community. Authors who earn the ire of the dominant clique can expect to have a harder time

> getting published and be quietly passed over at award ceremonies […]
>
> Brad R. Torgersen, who managed this year's Sad Puppies campaign, spoke to Breitbart London about its success: "I am glad to be overturning the applecart. Numerous authors, editors, and markets have been routinely snubbed or ignored over the years because they were not popular inside WSFS or because their politics have made them radioactive."
>
> Torgersen cites a host of authors who have suffered de facto exclusion from the sci-fi community: David Drake, David Weber, L.E Modesitt Jr, Kevn J. Anderson, Eric Flint, and of course Orson Scott Card — the creator of the world-famous Ender's Game, which was recently adapted into a successful movie. Despite his phenomenal success, Scott Card has been ostracized by sci-fi's inner circle thanks to his opposition to gay marriage.

I see minimal awareness from the social justice movement and the anti-social-justice movement that their narratives are similar, and certainly no deliberate intent to copy from one another. That makes me think of this as a case of convergent evolution.

The social justice attitude evolved among minority groups living under the domination of a different culture, which at best wanted to ignore them and at worst actively loathed them for who they were and tried to bully them into submission. The closest the average white guy gets to that kind of environment is wandering into a social-justice-dominated space and getting to experience the same casual hatred and denigration for them and everyone like them, followed by the

same insistence that they're imagining things and how dare they make that accusation and actually everything is peachy.

And maybe that very specific situation breeds a very specific kind of malignant hypervigilance, sort of halfway between post-traumatic stress disorder and outright paranoia, which motivates the obvious fear and hatred felt by both groups.

Someone is going to freak out and say I am a disgusting privileged shitlord for daring to compare the experience of people concerned about social justice to the experience of genuinely oppressed people, but they really shouldn't. That's the *explicit goal* of large parts of the social justice movement. For example, on the Hacker News thread about far-rightist Curtis Yarvin being kicked out of a tech conference for his views, one commenter writes:

> I've been involved in anti-racist/anti-fascist work, either directly or on the periphery, for about ten years at this point. This takes many forms, from street confrontations with fascists, protests at book readings and other events, and also disrupting fascist conferences and similar […]
>
> As far as this issue and other similar issues are concerned, I'm overjoyed that, as you put it, a climate of fear exists for fascists, misogynists, racists, and similar. I hope that this continues and only worsens for these people.
>
> I'm happy for many reasons. The first is that it has, as you've said, made privileged people afraid. I think this is only the beginning. Privilege creates safety, and as it is removed, I think the unsafety of the oppressed will in part come to the currently privileged classes. But if I could flip a switch and make every man feel the persistent, gnawing fear that a woman has of men, I would in a

> heartbeat. I wouldn't even consider whether the consequences were strategic, I would just do it.

This not the only time I've heard this opinion expressed, just the most recent. I feel like if you admit that you're trying your hardest to make privileged people feel afraid and uncomfortable and under siege in a way much like minorities traditionally do, and privileged people are in fact complaining of feeling afraid and uncomfortable and under siege in a way much like minorities traditionally do, you shouldn't immediately doubt their experience. Give yourself some more credit than that. You've been working hard, and at least in a few isolated cases here and there it's paid off.

The commenter continues:

> I would not say that I set out to defeat a "discourse-stifling" monster. The monsters I set out to defeat were patriarchy, capitalism, and white supremacy. These systems violently oppress, they don't "stifle discourse." In fact, they LOVE discourse! When people are discoursing, they aren't in the streets. I've seen so many promising movements hobbled by reformism that I'm glad the possibility no longer exists, though that isn't at all the fault of SJW-outrage (and is rather a consequence of the fact that the economy is in large part so perilous that nobody can afford the concessions that were previously won by reformists). So if discourse is permanently removed as a tactical and strategic option for future leftists, I'll consider it a victory.

Needless to say, that is not this blog's philosophy. But I think there is nevertheless something to be gained from all of the

hard work this guy and his colleagues have put in making other people feel unsafe.

The mirror neuron has always been one of liberalism's strongest weapon. A Christian doesn't decide to tolerate Islam because she likes Islam, she decides to tolerate Islam because she can put herself in a Muslim's shoes and realize that banning Islam would make him deeply upset in the same way that banning Christianity would make *her* deeply upset.

If the fear and hypervigilance that majority groups feel in social-justice-dominated spaces is the same as the fear and hypervigilance that minority groups feel in potentially discriminatory spaces, that gives us a whole lot more mirror neurons to work with and allows us to get a gut-level understanding of the other side of the dynamic. It lets us check my intuitions against their own evil twins on the other side to determine when we are proving too much.

II.

A couple of months ago the owners of a pizzeria mentioned in an interview that they wouldn't serve pizza at gay weddings because they're against gay marriage. Instantly the nation united in hatred of them and sent a bunch of death threats and rape threats and eventually they had to close down.

I thought this was ridiculous. I mean, obviously death threats are never acceptable, but there seemed to be something especially frivolous about this case, where there are dozens of other pizzerias gay people can go to and where *no one would ever serve pizza at a wedding anyway*. A pizzeria hardly holds the World Levers Of Power, so just let them have their weird opinion. All they're doing is sending potential paying customers to their more tolerant competitors, who are laughing all the way to the bank. It's a self-punishing offense.

This was very reasonable of me and I should be praised for my reasonableness, *except* that when a technology conference recently booted a speaker for having far-right views on his own time, I was one of the many people who found this really scary and thought they needed to be publicly condemned for this intolerant act.

In theory, the same considerations ought to apply. There are dozens of other technology conferences in the world. Technology conferences *also* do not hold the World Levers Of Power. And when they reject qualified rightist speakers, that just means they're just making life easier for their competitors who will be happy to grab the opportunity and laugh all the way to the bank. It ought to be self-punishing, so what's the worry.

My brain is *totally not on board* with this reasoning. When I ask it why, it says something like "No, you don't understand, these people are relentless, unless they are constantly pushed against they will put pressure on more and more institutions until their enemies are starved out or limited to tiny ghettos. Then they will gradually expand the definition of 'enemy' until everybody who doesn't do whatever they say is blacklisted from everywhere."

And if you think that's hyper-paranoid, then, well, you're probably right, but at least I have a lot of company. Here are some other comments on the same situation from the last links thread:

> I spent a semester of college in Massachusetts. That's where I found out that there are a lot of people who'd kill me and most of my family if they were given the chance. And thought it was totally reasonable and acceptable to say as much. (The things that are associated with Tumblr

these days existed long before it. And mostly came from academia.)

About the same time that sort of thing was happening in that online community, the same thing was happening in the real-world meat-space gatherings, also quite literally with shrill screams, mostly by [reacted] [reacted]s, who would overhear someone else's private conversations, and then start streaming "I BEG YOUR PARDON!" and "HOW CAN YOU SAY THAT!", and by [reacted] [reacted]'s who were bullying their way onto programming committees, and then making sure that various speakers, panelists, artists, authors, dealers, and GoHs known to be guilty of wrongthink were never invited in the first place. Were it not for the lucky circumstance of the rise of the web, the market takeoff of ebooks, especially a large ebook vendor (named after a river)'s ebook direct program, and the brave anchoring of a well known genre publisher that was specifically not homed in NYC, the purging of the genre and the community would have been complete.

Almost nobody wants to physically murder and maim the enemy, at least at the start. That's, well, the Final Solution. Plan A is pretty much always for the enemy to admit their wrongness or at least weakness, surrender, and agree to live according to the conqueror's rules. Maybe the leaders will have to go to prison for a while, but everyone else can just quietly recant and submit, nobody has to be maimed or killed. [The social justice community] almost certainly imagine they can achieve this through organized ostracism, social harassment, and democratic political activism. It's when they find that this

> won't actually make all the racists shut up and go away, that we get to see what their Plan B, and ultimately their final solution, look like.

And if you think my commenters are also hyper-paranoid, then you're probably *still* right. But it seems like the same kind of paranoia that makes gay people and their allies scream bloody murder against a single pizzeria, the kind that makes them think of it as a potential existential threat even though they've won victory after victory after victory and the only question still in the Overton Window is the terms of their enemies' surrender.

I mocked the hell out of the people boycotting Indiana businesses because of their right-to-discriminate law:

> Can we admit it's KIND OF funny ppl are boycotting Indiana for the immoral act of allowing people to boycott those they think act immorally?
>
> — Scott Alexander (@slatestarcodex) March 31, 2015

But if some state were to pass a law specifically saying "It is definitely super legal to discriminate against conservatives for their political beliefs," this would *freak me out*, even though I am not conservative and *even though this is already totally legal so the law would change nothing*. I would not want to rule out any response, up to and including salting their fields to make sure no bad ideas could ever grow there again.

Like many people, I am not very good at consistency.

III.

Author John Green writes books related to social justice. A couple of days ago, some social justice bloggers who disagreed with his perspective decided that a proportional

response was to imply he was a creep who might sexually abuse children. Green was somewhat put out by this, and said on his Tumblr that he was "tired of seeing the language of social justice – important language doing important work – misused as a way to dehumanize others and treat them hatefully" and that he thought his harassers "were not treating him like a person".

Speaking of the language of social justice, "dehumanizing" and "not treating like a person" are some pretty strong terms. They're terms I've criticized before – like when feminists say they feel like women aren't being treated as people, I'm tempted to say something like "the worst you've ever been able to find is a single-digit pay gap which may or may not exist, and you're going to turn that into people not thinking you're human?"

Here's another strong term: "hatred". The activist who got Mencius Moldbug banned from Strange Loop reassured us that he would never want someone banned merely for having unusual political views, but Moldbug went beyond that into "hatred", which means his speech is "hate speech", which is of course intolerable. This is a *bit* strange to anybody who's read any of his essays, which seem to have trouble with any emotion beyond smugness. I call him a bloodless and analytical thinker; the idea of his veins suddenly bulging out when he thinks about black people is too silly to even talk about. The same is true of the idea that people should feel "unsafe" around him; his entire shtick is that no one except the state should be able to initiate violence!

Likewise, when people wanted TV star Phil Robertson fired for saying (on his own time) that homosexuality was unnatural and led to bestiality and adultery, they said it wasn't about policing his religion, it was about how these were "hateful"

comments that would make the people working with him feel unsafe. At the time I said that was poppycock and that people who wanted him fired for having a private opinion were the worst kinds of illiberal witch-hunters.

On the other hand, consider Irene Gallo. I know nothing of her except what the Alas blog post says, but apparently in science fiction's ongoing conflict between the establishment and the anti-SJW "Sad Puppies"/"Rabid Puppies" groups, she referred to the latter as:

> Two extreme right-wing to neo-nazi groups that are calling for the end of social justice in science fiction and fantasy. They are unrepentantly racist, sexist and homophobic.

These are some pretty strong allegations, and range from "false" to "bizarre"; Brad Torgenson, leader of the group she called "extreme right wing neo nazi unrepentant racists", is happily married to a black woman. And the people she's talking about are her company's authors and customers, which hardly seems like good business practice. Some authors have said they feel uncomfortable working for a company whose employees think of them that way, and others have suggested boycotting Tor until they make her apologize or fire her.

Barry says that since she said these on her own private Facebook page, it is a private opinion that it would be pretty censorious to fire her over. Part of me agrees.

On the other hand, if I were a sci-fi author in one of the groups that she was talking about, I'm not sure I'd be able to work with her. Like, really? You want me to sit across a table and smile at the woman who thinks I'm a racist sexist homophobic extremist neo-Nazi just because I disagree with her?

Robertson's comment is just standard having-theological-opinions. Like, "Christian thinks homosexuality is sinful, more at eleven." Big deal. But Gallo's comment feels more like white hot burning hatred. She's clearly too genteel to personally kill me, but one gets the clear impression that if she could just press a button and have me die screaming, she'd do it with a smile on her face.

But this is just interpretation. Maybe Gallo doesn't consider "neo-Nazi" a term of abuse. Maybe this was just her dispassionate way of describing a political philosophy with the most appropriate analogy she could think of.

It doesn't seem likely to me. Then again, even though it seems obvious to me that stating "homosexuality is sinful and similar to bestiality" is a theological position totally compatible with being able to love the sinner and hate the sin, gay people have a lot of trouble believing it. And although I cannot condone firing people for their private opinions, back when people were trying to get rid of Gawker honcho Sam Biddle for saying that "nerds should be constantly shamed and degraded into submission", God help me it certainly crossed my head that there were even the slightest consequences for this kind of behavior, maybe other social justice writers would stop saying and acting upon statements like that *all the frickin' time?*

Once again, I'm not scoring very highly in consistency here.

IV.

A little while ago I had a bad couple of days. Some people were suggesting I was a liability to a group I was part of because I'd written some posts critical of feminism, and I got in a big fight about it. Then someone sent my ex-girlfriend a Tumblr message asking if they'd broken up with me "because I was racist". Then despite my best efforts to prevent this, my

Facebook feed decided to show me a bunch of Gawker-style articles about “Are all white people to blame for [latest atrocity]? I was too exhausted to write a real blog post, so I just threw together a links post. Because among two dozen or so links there was one (1) to the Moldbug story previously mentioned above, one commenter wrote that “your links posts are becoming indistinguishable from Chaos Patch” (Chaos Patch is the links post of notable far-right blog Xenosystems).

So I decided to ban that commenter. But since I have a policy in place of waiting an hour before doing anything rash, I took a long walk, thought about it a bit, and settled for just yelling at him instead.

Is banning someone for a kind of meaningless barb excessive? Well, yes. But given everything else that had happened, I didn’t have the energy to deal with it, and since this is my blog and the one corner of the world I have at least a tiny bit of control over I could at least symbolically get rid of a small fraction of my problems.

Plus, to me the barb seemed like an obvious veiled threat. “As long as you post any links about rightist causes, I can accuse you of being far-right. And we all know what happens to far-right people, eh?”

So even though out of context it was about the most minimal hostility possible, barely rising to the level where somebody would say it was even capable of being a problem at all, in context it really bothered me and made me at least somewhat justifiably feel unsafe.

Ever since I learned the word “microaggression” I have been unironically fond of it.

> Microagressions. Nanoagressions. Picoagressions. The Planck Hostility.
>
> — Map of Territory (@MapOfTerritory) January 28, 2015

When I'm putting up with too much and I've used up my entire mental buffer, then somebody bothering me and hiding under the cover of "oh, this was such a tiny insult that you would seem completely crazy to call me on it" is *especially* infuriating, even more infuriating than someone insulting me outright and me being able to respond freely. The more you have to deal with people who hate you and want to exclude you, the more likely you are to get into this mode, not to mention people who have developed their own little secret language of insults.

Here's an example of what I mean by "secret language of insults": consider the term "dude", as in "white dude". There is nothing objectively wrong with "dude" when it is applied to surfers or something. But when a feminist says it, as in the term "white dudes", you know it is going to be followed by some claim that as a white dude, you are exactly the same as all other white dudes and entirely to blame for something you don't endorse. The first page of Google results is overratedwhitedudes.tumblr.com, Gawker saying Wimpy White Dudes Ruined American Idol, and Mother Jones saying glowingly that You Won't Find Many White Dudes At This Tech Startup. Being called a "white dude" is always followed by the implication that you're ruining something or that your very presence is cringeworthy and disgusting.

I had a feminist friend who used to use the term "dudes" for "men" all the time. I asked them to please stop. They said that was silly, because that was just the word the culture they'd

grown up in used, and obviously no harm was meant by it, and if I took it as an insult then I was just being oversensitive. This is *word for word* the explanation I got when I asked one of my elderly patients to stop calling black people *their* particular ethnic slur.

The counterpart to subliminal insults is superliminal insults; ones that are hard to detect because they're so over-the-top obvious.

I was recently reading a social justice blog where someone complained about men telling women "Make me a sandwich!" in what was obvious jest.

On the one hand, no one can possibly take this seriously.

On the other hand, there's a common social justice meme where people post under the hashtag #killallwhitemen.

Certainly this cannot be taken seriously; most social justice activists don't have the means to kill all white men, and probably there are several of them who wouldn't do it even if they could. It should not be taken, literally, as a suggestion that all white men should be killed. On the other hand, *for some bizarre reason* this tends to make white men uncomfortable.

The obvious answer is that the people posting "Wimmen, make me a sandwich!" don't literally believe that women exist only for making them sandwiches, but they *might* believe a much weaker claim along the same lines, and by making the absurd sandwich claim, they can rub it in while also claiming to be joking. At least this is how I feel about the "kill all white men" claim.

As long as you've got a secret language of insults that your target knows perfectly well are insulting, but which you can credibly claim are not insulting at all – maybe even believing

it yourself – then you have the ability to make them feel vaguely uncomfortable and disliked everywhere you go without even trying. If they bring it up, you can just laugh about how silly it is that people believe in "microaggressions" and make some bon mot about "the Planck hostility".

V.

I'm taking a pretty heavy Outside View line here, so let me allow my lizard brain a few words in its own defense.

"Yes," my lizard brain says, "social justice activists and the people silenced by social justice activists use some of the same terms and have some of the same worries. But the latter group has *reasonable* worries, and the former group has totally *unreasonable* worries, which breaks the symmetry."

Interesting. Please continue, lizard brain.

"Black people might be very worried about being discriminated against. But the chance that someone would say 'Let's ban all black people from our technology conference, because they are gross', and everyone would say 'Yes, that is a splendid idea', and the government and media would say 'Oh, wonderful, we are so proud of you for banning all black people from your conference' is zero point zero zero zero. On the other hand, this is something that conservatives worry about every day. The chance that someone would say 'You know, there's no reason raping women should be illegal, let's not even bother recording it in our official statistics' is *even lower than that*, but this is exactly what several countries do with male rape victims. If someone says 'kill all white men', then all we do is hold an interminable debate about whether that disqualifies them from the position of Diversity Officer; if someone said 'kill all gays', we would be much more final in pronouncing them Not Quite Diversity Officer Material."

But don't you –

"The reason why we don't care about a pizzeria that won't serve gay people is that recent years have shown an overwhelming trend in favor of more and more rights and acceptance of gay people, and the pizzeria is a tiny deviation from the pattern which is obviously going to get crushed under the weight of history even without our help. The reason we worry about a conference banning conservatives is that conservatives are an actually-at-risk group, and their exclusion could grow and grow until it reaches horrific proportions. The idea of a pizzeria banning gays and a conference banning conservatives may seem superficially similar out of context, but when you add this piece of context they're two completely different beasts."

Two responses come to mind.

First, this is obviously true and correct.

Second, this is exactly symmetrical to my least favorite argument, the argument from privilege.

The argument from privilege is something like "Yeah, sure, every so often the system is unfair to white people or men or whatever in some way. But this is not a problem and we should not even be talking about it, because privilege. Shows that mock women for stereotypically female failings are sexist, but shows that mock men for stereotypically male failings are hilarious, and you may not call them sexist because you can't be sexist against privileged groups."

My argument has always been "What's good for the goose is good for the gander".

But either this argument goes, or my lizard brain's argument goes, or we have to move to the object level, or somebody has

to get more subtle.

VI.

My point is, there are a lot of social justice arguments I *really* hate, but which I find myself unintentionally reinventing any time things go really bad for me, or I feel like myself or my friends are being persecuted.

I should stop to clarify something. "Persecuted" is a strong word. "Feel like we are being persecuted" is way weaker.

A couple weeks ago there was a Vox article, America's Never Been Safer, So Why Do Republicans Believe It Is In Mortal Peril?. It brought up a lot of cute statistics, like that the rate of pedestrians being killed by car accidents is much higher than the rate of civilians being killed in terrorist attacks. It joked that "You're over 100 times more likely to die by literally walking around than you are to be killed in a terrorist attack."

On the other hand, vox has practically led the news media in 24-7 coverage of police officers shooting unarmed black people, talking about how it's a huge threat to our values as a civilization and how white people don't understand that all black people have to constantly live in fear for their lives.

But a quick calculation demonstrates that unarmed black people are about 10 times more likely to die by *literally walking around* than by getting shot by a white police officer. One gets the feeling Vox doesn't find this one nearly as funny.

But here I would perform another quick calculation. Here's a list of people who have been publicly shamed or fired for having politically incorrect opinions. Even if we assume the list is understating the extent of the problem by an entire order of magnitude, you're *still* more likely to die by literally

walking around than you are to get purged for your politically incorrect opinion.

Like a lot non-feminists, I was freaked out by the recent story about a man who was raped while unconscious being declared the rapist and expelled from college without getting to tell his side of the story. I have no evidence that this has ever happened more than just the one time mentioned in the article, let alone it being a national epidemic that might one day catch me in its clutches, but because I've had to deal with overly feminist colleges in other ways, my brain immediately raised it to Threat Level Red and I had to resist the urge to tell my friends in colleges to get out while they still could. If we non-feminists can get worried about this – and we can – we have less than no right to tell feminists they shouldn't *really* be worried about college rape because the real statistics are 1 in X and not 1 in Y like they claim.

Hopefully some readers are lucky enough never to have felt much personal concern about terrorism, police shootings, rape, rape accusations, or political correctness. But if you've worried about at least one of these low-probability things, then I hope you can extend that concern to understand why other people might be worried about the others. It seems to have something to do with the chilling effect of knowing that something is intended to send a message to you, and in fact receiving that message.

(as an aside, I find it surprising that so many people, including myself, are able to accept the statistics about terrorism so calmly without feeling personally threatened. My guess is that, as per Part VIII here, we don't primarily identify as Americans, so a threat deliberately framed as wanting to make Americans feel unsafe just bounces off us.)

In an age where the media faithfully relates and signal-boosts all threats aimed at different groups, and commentators then serve their own political needs by shouting at us that WE ARE NOT FEELING THREATENED ENOUGH and WE NEED TO FEEL MORE THREATENED, it is very easy for a group that faces even a small amount of concerted opposition, even when most of society is their nominal allies and trying hard to protect them, to get pushed into a total paranoia that a vast conspiracy is after them and they will never be safe. This is obviously the state that my commenters who I quoted in Part II are stuck in, obviously the state that those people boycotting the Indiana pizzeria are stuck in, and, I admit, a state I'm stuck in a lot of the time as well.

VII.

Getting back to the thesis, my point is there are a lot of social justice arguments I *really* hate, but which I find myself unintentionally reinventing any time things go really bad for me, or I feel like myself or my friends are being persecuted.

Once events provoke a certain level of hypervigilance in someone – which is very easy and requires only a couple of people being hostile, plus the implication that they there's much more hostility hidden under the surface – then that person gets in fear for their life and livelihood and starts saying apparently bizarre things: that nobody treats them as a person, that their very right to exist is being challenged. Their increasingly strident rhetoric attracts increasingly strident and personal counter-rhetoric from the other side, making them more and more threatened until they reach the point where Israel is stealing their shoe. And because they feel like every short-term battle is the last step on the slippery slope to their total marginalization, they engage in crisis-mode short-term thinking and are understandably willing to throw longer-term

values like free speech, politeness, nonviolence, et cetera, under the bus.

Although it's very easy enter this state of hypervigilance yourself no matter how safe you are, it's very hard to understand why anyone else could possibly be pushed into it despite by-the-numbers safety. As a result, we constantly end up with two sides both shouting "You're making me live in fear, and also you're making the obviously false claim that you live in fear yourself! Stop it!" and no one getting anywhere. At worst, it degenerates into people saying "These people are falsely accusing me of persecuting them, *and* falsely claiming to be persecuted themselves, I'll get back at them by mocking them relentlessly, doxxing them, and trying to make them miserable!" and then you get the kind of atmosphere you find in places like SRS and Gamergate and FreeThoughtBlogs.

But I'm also slightly optimistic for the future. The conservative side seems to have been about ten years behind the progressive side in this, but they're catching up quickly. Now *everybody* has to worry about being triggered, *everybody* has to worry about their comments being taken out of context by Gawker/Breitbart and used to get them fired and discredit their entire identity group, *everybody* has to worry about getting death threats, et cetera. This is bad, but also sort of good. When one side has nukes, they nuke Hiroshima and win handily. When both sides have nukes, then under the threat of mutually assured destruction they eventually come up with protocols to prevent those nukes from being used.

Now that it's easier to offend straight white men, hopefully they'll agree trigger warnings can be a useful concept. And now that some social justice activists are getting fired for voicing their opinions in private, hopefully they'll agree that you shouldn't fire people for things they say on their own

time. Once everyone agrees with each other, there's a chance of getting somewhere. Yes, all of this will run up against a wall of "how dare you compare what I'm doing to what you're doing, I'm defending my right to exist but you're engaging in hate speech!" but maybe as everyone gets tired of the nukes flying all the time people will become less invested in this point and willing to go to the hypothetical Platonic negotiation table.

My advice for people on the anti-social justice side – I don't expect giving the SJ people advice would go very well – is that it's time to stop talking about how social justice activism is necessarily a plot to get more political power, or steal resources, or silence dissenting views. Like everything else in the world it can certainly turn into that, but I think our *own* experience gives us a lot of reasons to believe they're exactly as terrified as they say, and that we can't expect them to accept "you have no provable objective right to be terrified" any more than our lizard brains would accept it of us. I think it's time to stop believing that they censor and doxx and fire their opponents out of some innate inability to understand liberalism, and admit that they probably censor and doxx and fire their opponents because they're as scared as we are and feel a need to strike back.

This isn't a claim that they don't have it in for us – many of them freely admit they do – and that they don't need to be stopped. It's just a claim that we can gain a good understanding of *why* they have it in for us, and how we might engineer stopping them in a way less confrontational than fighting an endless feud.

Yesterday, a friend on Facebook posted something about a thing men do which makes women feel uncomfortable and which she wanted men to stop. I carefully thought about

whether I ever did it, couldn't think of a time I had, but decided to make sure I didn't do it in the future.

I realized that if I'd heard the exact same statement from Gawker, I would have interpreted it (correctly) as yet another way to paint men as constant oppressors and women as constant victims in order to discredit men's opinions on everything, and blocked the person who mentioned it to me so I didn't have to deal with yet another person shouting that message at me. The difference this time was that it came from an acquaintance who was no friend of feminism, who has some opinions of her own that might get her banned from tech conferences, and who I know would have been equally willing to share something women do that bothers men, if she had thought it important.

If we can get to a point where we don't feel like requests are part of a giant conspiracy to discredit and silence us, people *are* sometimes willing to listen. Even *me*.

Archipelago and Atomic Communitarianism

I.

Forty years ago, Robert Nozick proposed a very strange utopia, which he considered the culmination of libertarian principles.

Ten years ago, Mencius Moldbug proposed the same utopia, considering it the culmination of conservative principles.

Three years ago, unaware of either, I independently invented a role-playing game around the same utopia, considering it the culmination of liberal principles.

Nozick called it Meta-Utopia. Moldbug called it Patchwork. I called it Archipelago.

II.

In 2011 the conworlding experiment I'd been part of for the past ten years, Micras, was starting to wind down. There were lots of reasons, but a big one was the trouble getting everyone working together on a coherent world. Some participants wanted to simulate medieval countries with wizards and dragons. Other people were more into modern nation-states with factories and steel production quotas. Still others wanted to do scifi stuff where they debated the ethics of genetic engineering and maybe built mile-long starships.

All of which were fine, *until* you tried to stick it together into a coherent world, at which point it made no sense. How come the people with mile-long starships hadn't invaded the people who were still jousting on horseback? How come the industrialists could spend two hundred years worrying about

steel quotas without inventing any new technology beyond steel, let alone stealing or buying the technology from the scifi civilization next door? If magic worked, how come only one or two civilizations were using it?

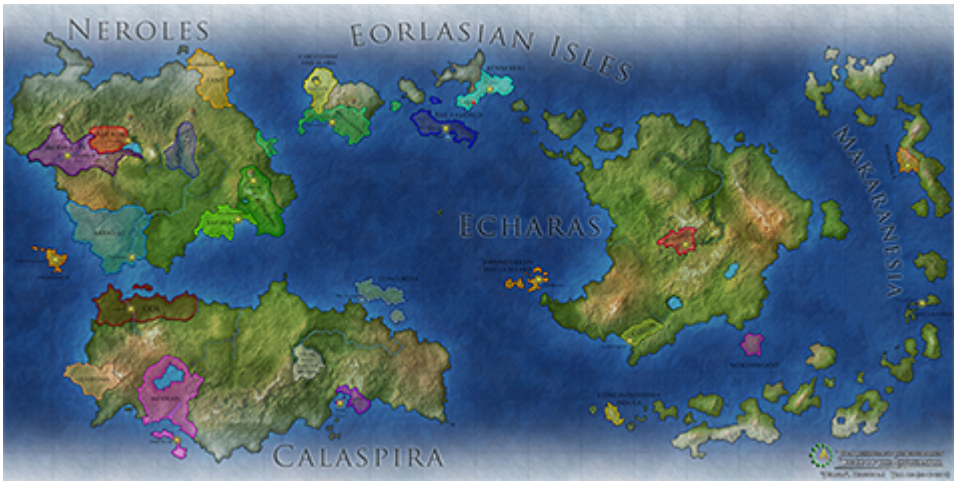
Solving these problems tended to involve a *lot* of just-so stories, but the more we came up with, the harder it was to support any kind of interaction at all.

I decided to sweep the whole thing under the carpet with a parallel-running Gritty Reboot. I dug out my old transhuman goddess character Maria Morimoto and had her wipe out civilization, leaving only scattered barbarians and a few groups who had managed to keep the vestiges of civilization together. Her human viceroy, Omi Oitherion, gathered the last remnants of civilization and forged them into a world government called Archipelago whose goal was to create utopia through a process of evolution and experimentation.

The way it worked was that any tribe of surviving civilized humans with a coherent philosophy could apply to become part of Archipelago. Their application would include the location of their desired homeland, and the technological and magical level they found most conducive to human flourishing. Archipelago would then use its transhuman powers to trap their homeland in a telluric field limiting it to exactly that level of technology + magic, and protect them from incursion by any other group.

At first, the conceit worked really well. I granted myself absolute power. Alicorn got elected as the democratic figurehead. The online infrastructure got set up. About two dozen conworlders agreed to participate. New "statelets" sprung up just about weekly. We got everything from inoffensive liberal democracies to monastic religious orders to

socialist communes to tribes of violent cannibals to one guy who tried to recreate Plato's Republic in a giant underground cave.



Pictured: Pelagia, the game world of Archipelago. Click to enlarge.

After about a year or so, it started working less well. By bad luck, a couple of the major players left all at once. Activity died down. Micras, still running in parallel, started doing better and the need for an alternative became less pressing. I tried to shape the backstory and more than some people were comfortable with. Whatever. Archipelago went quiet, we switched off the lights, and activity shifted back to Micras. It became one of those pieces of legend you get in all Internet communities: "Remember that time we tried something kind of cool, but it didn't work out?"

Freed from the responsibility of running a real game, I started writing more of the story of Archipelago in my head, fleshing out details. Why try to maximize diversity of cultural experiments? What was Omi's endgame? What precisely were the bylaws of the World Government?

Gradually, some ideas started to take shape.

III.

In the old days, you had your Culture, and that was that. Your Culture told you lots of stuff about what you were and weren't allowed to do, and by golly you listened. Your Culture told you to work the job prescribed to you by your caste and gender, to marry who your parents told you to marry or at *least* someone of the opposite sex, to worship at the proper temples and the proper times, and to talk about *proper* things as opposed to the blasphemous things said by the tribe over there.

Then we got Liberalism, which said all of that was mostly bunk. Like Wicca, its motto is "Do as you will, so long as it harms none". Or in more political terms, "Your right to swing your fist ends where my nose begins" or "If you don't like gay sex, don't have any" or "If you don't like this TV program, don't watch it" or "What happens in the bedroom between consenting adults is none of your business" or "It neither breaks my arm nor picks my pocket". Your job isn't to enforce your conception of virtue upon everyone to build the Virtuous Society, it's to live your own life the way you want to live it and let other people live *their* own lives the way *they* want to live them.

This is the much-maligned "atomic individualism", or at least one definition of such. I'm not sure anyone has a great idea what it means; it seems to be more a bogeyman for conservatives to take potshots at than a position with its own supporters and think tanks or anything.

On the other hand, the Left is starting to get pretty wary of atomic individualism too. Maybe one of the first signs of the trend was tobacco ads. Even though putting up a billboard saying "SMOKE MARLBORO" neither breaks anyone's arm nor picks their pocket, it shifts social expectations in such a way that bad effects occur. It's hard to dismiss that with "Well, it's people's own choice to smoke and they should live their

lives the way they want" if studies show that more people will want to live their lives in a way that gives them cancer in the presence of the billboard than otherwise.

From there we go into policies like Michael Bloomberg's ban on giant sodas. While the soda ban itself was probably as much symbolic as anything, it's hard to argue with the impetus behind it – a culture where everyone gets exposed to the option to buy very very unhealthy food all the time is going to be less healthy than one where there are some regulations in place to make EAT THIS DONUT NOW a less salient option. I mean, I *know* this is true. A few months ago when I was on a diet I *cringed* every time one my coworkers would bring in a box of free donuts and place it in, wide-open, in the doctors' lounge, because there was *no way* I wasn't going to take one (or two, or three). I could ask people to stop, but they probably wouldn't, and then it would be a *different* place where I encounter the wide-open box of free donuts. I am not proposing that it is *ethically wrong* to bring in free donuts or that banning them is the correct policy, but I do want to make it clear that stating "it's your free choice to partake or not" doesn't eliminate the problem, and that this points to an entire class of serious issues where Liberalism as construed above is at best an imperfect heuristic.

And I would be remiss talking about the left's turn away from Liberalism without mentioning social justice. The same people who once deployed Liberal arguments against conservatives: "If you don't like profanity, don't use it", "If you don't like this offensive TV show, don't watch it", "If you don't like pornography, don't buy it" – are now concerned about people using ethnic slurs, TV shows without enough minority characters, and pornography that encourages the objectification of women. I've objected to some of this on

purely empirical grounds, but the least convenient possible world is the one where the purely empirical objections fall flat. If they ever discover proof positive that yeah, pornographication makes women hella objectified, is it acceptable to censor or ban misogynist media on a society-wide level?

And if the answer is yes – and if such media like really, *really* increases the incidence of rape I'm not sure how it couldn't be – then what about all those conservative ideas we've been neglecting for so long? What if strong, cohesive, religious, demographically uniform communities make people more trusting, generous, and cooperative in a way that *also* decreases violent crime and other forms of misery? We have some good evidence lots of evidence that this is true, and although we can doubt each individual study, we owe conservatives the courtesy of imagining the possible world in which they are right, the same as anti-misogyny leftists. Maybe media glorifying criminals or lionizing nonconformists above those who quietly follow cultural norms has the same kind of erosive effects on "values" as misogynist media. Or, at the very least, we ought to have a good philosophy in place so that we have some idea what to do it if does.

IV.

A while ago, in Part III of this essay, I praised liberalism as the only peaceful answer to Hobbes' dilemma of the war of all against all.

Hobbes' point, remember, is that if everyone's fighting everyone loses out. Even the winners probably end up worse off than if they had just been able to live in peace. He says that governments are good ways to prevent this kind of conflict. Someone – in his formulation a king – tells everyone else what

they're going to do, and then everyone else does it. No fighting necessary. If someone tries to start a conflict by ignoring the king, the king quashes them with such overwhelming force that it doesn't even count as a fight.

But this replaces the problem of potential warfare with the problem of potential tyranny. So we've mostly shifted from absolute monarchies to other forms of government, which is all nice and well except that governments allow a *different* kind of war of all against all. Instead of trying to kill their enemies and steal their stuff, people are tempted to ban their enemies and confiscate their stuff. Instead of killing the Protestants, the Catholics simply ban Protestantism. Instead of forming vigilante mobs to stone homosexuals, the straights merely declare homosexuality is punishable by death. It *might* be better than the alternative – at least everyone knows where they stand and things stay peaceful – but the end result is still a lot of pretty miserable people.

Liberalism is a new form of Hobbesian equilibrium where the government enforces not only a ban on killing and stealing from people you don't like, but also a ban on tyrannizing them out of existence. This is the famous "freedom of religion" and "freedom of speech" and so on, as well as the "freedom of what happens in the bedroom between consenting adults". The Catholics don't try to ban Protestantism, the Protestants don't try to ban Catholicism, and everyone is happy.

Liberalism only works when it's clear to everyone on all sides that there's a certain neutral principle everyone has to stick to. The neutral principle can't be the Bible, or Atlas Shrugged, or anything that makes it look like one philosophy is allowed to judge the others. Right now that principle is the Principle of Harm: you can do whatever you like unless it harms other people, in which case stop. We seem to have inelegantly

tacked on an "also, we can collect taxes and use them for a social safety net and occasional attempts at social progress", but it seems to be working pretty okay too.

The Strict Principle of Harm says that pretty much the only two things the government can get angry at is literally breaking your leg or picking your pocket – violence or theft. The Loose Principle of Harm says that the government can get angry at complicated indirect harms, things that Weaken The Moral Fabric Of Society. Like putting up tobacco ads. Or having really really big sodas. Or publishing hate speech against minorities. Or eroding trust in the community. Or media that objectifies women.

No one except the most ideologically pure libertarians seems to want to insist on the Strict Principle of Harm. But allowing the Loose Principle Of Harm restores all of the old wars to control other people that liberalism was supposed to prevent. The one person says "Gay marriage will result in homosexuality becoming more accepted, leading to increased rates of STDs! That's a harm! We must ban gay marriage!" Another says "Allowing people to send their children to non-public schools could lead to kids at religious schools that preach against gay people, causing those children to commit hate crimes when they grow up! That's a harm! We must ban non-public schools!" And so on, forever.

And I'm talking about non-governmental censorship just as much as government censorship. Even in the most anti-gay communities in the United States, the laws usually allow homosexuality or oppose it only in very weak, easily circumvented ways. The real problem for gays in these communities is the social pressure – whether that means disapproval or risk of violence – that they would likely face for coming out. This too is a violation of Liberalism, and it's

one that's as important or more important than the legal version.

And right now our way of dealing with these problems is to argue them. "Well, gay people don't really increase STDs too much." Or "Home-schooled kids do better than public-schooled kids, so we need to allow them." The problem is that arguments never terminate. Maybe if you're *incredibly* lucky, after years of fighting you can get a couple of people on the other side to admit your side is right, but this is a pretty hard process to trust. The great thing about religious freedom is that it short-circuits the debate of "Which religion is correct, Catholicism or Protestantism?" and allows people to tolerate both Catholics and Protestants even if they are divided about the answer to this object-level question. The great thing about freedom of speech is that it short-circuits the debate of "Which party is correct, the Democrats or Republicans?" and allows people to express both liberal and conservative opinions even if they are divided about the object-level question."

If we force all of our discussions about whether to ban gay marriage or allow home schooling to depend on resolving the dispute about whether they indirectly harm the Fabric of Society in some way, we're forcing dependence on object-level arguments in a way that historically has been very very bad.

Presumably here the more powerful groups would win out and be able to oppress the less powerful groups. We end up with exactly what Liberalism tried to avoid – a society where everyone is the guardian of the virtue of everyone else, and anyone who wants to live their lives in a way different from the community's consensus is out of luck.

In Part III, I argued that *not allowing* people to worry about culture and community at all was inadequate, because these things really do matter.

Here I'm saying that if we *do allow* people to worry about culture and community, we risk the bad old medieval days where all nonconformity gets ruthlessly quashed.

Right now we're balanced precariously between the two states. There's a lot of Liberalism, and people are generally still allowed to be gay or home-school their children or practice their religion or whatever. But there's also quite a bit of Enforced Virtue, where kids are forbidden to watch porn and certain kinds of media are censored and in some communities mentioning that you're an atheist will get you Dirty Looks.

It tends to work okay for most of the population. Better than the alternatives, maybe? But there's still a lot of the population that's not free to do things that are very important to them. And there's also a lot of the population that would like to live in more "virtuous" communities, whether it's to lose weight faster or avoid STDs or not have to worry about being objectified. Dealing with these two competing issues is a pretty big part of political philosophy and one that most people don't have any principled solution for.

V.

Here is where my meditations on Archipelago took me.

Imagine Dragumve, the only city-state to survive the apocalypse that destroyed Micras relatively intact. Tempered by years of surviving famine and disease and barbarian attack, it's a pretty grim place. But now the century-long winter has ended, the barbarians have mostly been pushed away behind natural borders, and things are looking up. Its inhabitants start to have some time to philosophize, and they all have some

different conceptions of the good life. They start fighting on what the political system of Dragumve should look like.

Omi Oitherion, the absolute ruler of Dragumve, says – Here, we're doing things my way. But those of you with different ideas, go forth and settle the world, and I won't stop you. In fact, I'll protect you. Go found city-states based on your philosophies.

And so the equivalent of our paleoconservatives go out and found communities based on virtue, where all sexual deviancy is banned and only wholesome films can be shown and people who burn the flag are thrown out to be eaten by wolves.

And the equivalent of our social justiciars go out and found communities where all movies have to have lots of strong minority characters in them, and all slurs are way beyond the pale, and nobody misgenders anybody.

And the equivalent of our Objectivists go out and found communities based totally on the Strict Principle of Harm where everyone is allowed to do whatever they want and there are no regulations on business and everything is super-capitalist all the time.

And some people who just really want to lose weight go out and found communities where you're not allowed to place open boxes of donuts in the doctors' lounge.

Usually the communities are based on a charter, which expresses some founding ideals and asks only the people who agree with those ideals to enter. The charter also specifies a system of government. It could be an absolute monarch, charged with enforcing those ideals upon a population too stupid to know what's good for them. Or it could be a direct democracy of people who all agree on some basic principles

but want to work out for themselves what direction the principles take them.

After a while, Omi Oitherion, who remember is the viceroy for a transhuman goddess and is kind of omnipotent, decides to formalize and strengthen this system, not to mention work out some of the ethical dilemmas.

The first thing he does is ban communities from declaring war on each other. That's an *obvious* gain. He could just smite warmongers, but he thinks it's more natural and organic to get all the communities into a World Government. Every community donates a certain amount to a military, and the military's only job is to quash anyone from any community who tries to invade another.

The second thing World Government does is address externalities. For example, if some communities emit a lot of carbon, and that causes global warming which threatens to destroy other communities, the World Government puts a stop to that. If the offending communities refuse to stop emitting carbon, then there's that military again.

The third thing World Government does is prevent memetic contamination. If one community wants to avoid all media that objectifies women, then no other community is allowed to broadcast women-objectifying media in. If a community wants to live an anarcho-primitivist lifestyle, nobody else is allowed to import TVs. Every community decides *exactly* how much informational contact it wants to have with the rest of the world, and no one is allowed to force them to have more than that.

The most important job for World Government is to think of the children.

Imagine you're conservative Christians, and you're tired of this secular godless world, so you go off with your conservative Christian friends to found a conservative Christian community. You all pray together and stuff and are really happy. Then you have a daughter. Turns out she's atheist and lesbian. What now?

Well, it might be that your kid would be much happier at the lesbian separatist community down the road. The *absolute minimum* that the World Government can do is enforce freedom of movement. That is, the *second* your daughter decides she doesn't want to be in Christiantopia anymore, she goes to a World Government embassy nearby and asks for a ticket out, which they give her, free of charge. She gets airlifted to Lesbiantopia the next day. If *anyone* in Christiantopia tries to prevent her from reaching that embassy, or threatens her family if she leaves, or expresses the *slightest* amount of coercion to keep her around, World Government *notices*.

Those of my readers who were involved in the Archipelago project may remember that Omi Oitherion is *not* a good person to offend.

But this is not nearly enough to fully solve the child problem. A child who is abused may be too young to know that escape is an option, or may be brainwashed into thinking they are evil, or guilted into believing they are betraying their families to opt out. And although there is no perfect, elegant solution here, the practical solution is that World Government enforces some pretty strict laws on child-rearing, and every child, no matter what other education they receive, also has to receive a class taught by a World Government representative in which they learn about the other communities participating in Archipelago, receive a basic non-brainwashed view of the

world, and are given directions to their nearest World Government representative who they can give their opt-out request to.

The list of communities they are informed about always starts with Dragumve, which is ruled by World Government itself and is considered an inoffensive, neutral option for people who don't want anywhere in particular. And it always ends with a reminder that if they can gather enough support, World Government will provide them with help for an expedition to go out and found their own community somewhere in the wilderness.

There's one more problem World Government has to deal with, which is malicious inter-community transfer. Suppose that there is some community which puts extreme effort into educating its children, an education which it supports through heavy taxation. New parents move to this community, reap the benefits, and then when their children grow up they move back to their previous community so they don't have to pay the taxes to educate anyone else. The communities themselves prevent some of this by immigration restrictions – anyone who's clearly taking advantage of them isn't allowed in (except in Dragumve, which has an official committment to let in anyone who wants). But that still leaves the example of people maliciously leaving a high-tax community once they've got theirs. I imagine this is a big deal in Archipelago politics, but that in practice World Government asks these people, even in their new homes, to pay higher tax rates to subsidize their old community. Or since that could be morally objectionable (imagine the lesbian separatist having to pay taxes to Christiantopia which oppressed her), maybe they pay the excess taxes to World Government itself, as a way of disincentivizing malicious movement.

Because there *are* World Government taxes, and most people are happy to pay them. In my fantasy, World Government isn't an enemy, where the Christians view it as this evil atheist conglomerate trying to steal their kids away from them and the capitalists view it as this evil socialist conglomerate trying to enforce high taxes. The Christians, the capitalists, and everyone else are extraordinarily *patriotic* about being part of the Archipelago, for its full name is the Archipelago of Civilized Communities, it is the standard-bearer of civilization against the barbarian hordes, and it is precisely the institution that allows them to maintain their distinctiveness in the face of what would otherwise be irresistable pressure to conform. Atheistopia is the enemy of Christiantopia, but only in the same way the Democratic Party is the enemy of the Republican Party – two groups within the same community who may have different ideas but who consider themselves part of the same broader whole, fundamentally allies under a banner of which both are proud.

The banner, by the way, looks a lot like the EU flag. I'm not sure how to feel about that.

VI.

It's easy to see why Robert Nozick thinks this is a libertarian utopia. World Government does very very little. Other than the part with children and the part with evening out taxation regimes, it just sits around preventing communities from using force against each other. That makes it very very easy for anyone who wants freedom to start a community that grants them the kind of freedom they want – or, more likely, to just start a community organized on purely libertarian principles. The World Government of Archipelago is the perfect minarchist night watchman state, and any additions you make over that are chosen by your own free will.

And it's easy to see why other people think this is a conservative utopia. Conservativism, when it's not just Libertarianism Lite, is about building strong cohesive communities of relatively similar people united around common values. Archipelago is obviously built to make this as easy as possible, and it's hard to imagine that there wouldn't pop up a bunch of communities built around the idea of Decent Small-Town God-Fearing People where everyone has white picket fences and goes to the same church and nobody has to lock their doors at night (so basically Utah; I feel like this is one of the rare cases where the US' mostly-in-name-only Archipelagoness really asserts itself). People who didn't fit in could go to a Community Of People Who Don't Fit In and would have no need to nor right to complain, and they wouldn't have to deal with Those Durned Bureaucrats In Washington telling them what to do.

But to me, this seems like a liberal utopia, even a leftist utopia, for three reasons.

The first reason is that it extends the basic principle of liberalism – solve differences of opinion by letting everyone do their own thing according to their own values, then celebrate the diversity this produces. I like homosexuality, you don't, fine, I can be homosexual and you don't have to, and having both gay and straight people living side by side enriches society. This just takes the whole thing one meta-level up – I want to live in a very sexually liberated community, you want to live in a community where sex is treated purely as a sacred act for the purpose of procreation, fine, I can live in the community I want and you can live in the community you want, and having both sexually-liberated and sexually-pure communities living side by side enriches

society. It is pretty much saying that the solution to any perceived problems of liberalism is *much more liberalism*.

The second reason is quite similar to the conservative reason. A lot of liberals have some pretty strong demands about the sorts of things they want society to do. I was recently talking to Ozy about a group who believe that society billing thin people is fatphobic, and that everyone needs to admit obese people can be just as attractive and date more of them, and that anyone who preferentially dates thinner people is Problematic. They also want people to stop talking about nutrition and exercise publicly. I sympathize with these people, especially having recently read a study showing that obese people are much happier when surrounded by other obese, rather than skinny people. But realistically, their movement will fail, and even philosophically, I'm not sure how to determine if they have the right to demand what they are demanding or what that question means. Their best bet is to found a community on these kinds of principles and only invite people who already share their preferences and aesthetics going in.

The third reason is the reason I specifically draw leftism in here. Liberalism, and to a much greater degree leftism, are marked by the emphasis they place on oppression. They're particularly marked by an emphasis on oppression being a really hard problem, and one that is structurally inherent to a certain society. They are marked by a moderate amount of despair that this oppression can ever be rooted out.

And I think a pretty strong response to this is making sure everyone is able to say "Hey, you better not oppress us, because if you do, we can pack up and go somewhere else."

Like if you want to protest that this is unfair, that people shouldn't be forced to leave their homes because of

oppression, fine, fair enough. But given that oppression *is* going on, and you haven't been able to fix it, giving people the *choice* to get away from it seems like a pretty big win. I am reminded of the many Jews who moved from Eastern Europe to America, the many blacks who moved from the southern US to the northern US or Canada, and the many gays who make it out of extremely homophobic areas to friendlier large cities. One could even make a metaphor, I think rightly, to telling battered women that they are allowed to leave their husbands, telling them they're not forced to stay in a relationship that they consider abusive, and making sure that there are shelters available to receive them.

If any person who feels oppressed can leave whenever they like, to the point of being provided a free plane ticket by the government, how long can oppression go on before the oppressors give up and say "Yeah, guess we need someone to work at these factories now that all our workers have gone to the communally-owned factory down the road, we should probably at least let people unionize or something so they will tolerate us"?

This is pretty funny, because the idea I'm pushing is rather explicitly reactionary. Like, I think it would be fair to call this the *single core idea* of reaction. All that stuff about kings and gender roles and ethno-nationalism is to some degree idle speculation about what kind of Archipelagian community would end up most successful, in the same way transhumanists sometimes speculate about how things should be run after the Singularity.

Yet I think its liberal credentials are impeccable. A commenter in the latest Asch thread mentioned an interesting quote by Frederick Douglass:

> The American people have always been anxious to know what they shall do with us [black people]. I have had but one answer from the beginning. Do nothing with us! Your doing with us has already played the mischief with us. Do nothing with us!

It sounds like, if Frederick Douglass had the opportunity to go to some other community, or even found a black ex-slave community, no racists allowed, he probably would have taken it [edit: or not, or had strict conditions]. If the people in slavery during his own time period had had the chance to leave their plantations for that community, I bet they would have taken it too. And if you believe there are still people today whose relationship with society are similar in kind, if not in degree, to that of a plantation slave, you should be pretty enthusiastic about the ability of exit rights and free association to disrupt those oppressive relationships.

VII.

We lack Archipelago's big advantage – a vast frontier of unsettled land.

Which is not to say that people don't form communes. They do. Some people even have really clever ideas along these lines, like the seasteaders. But the United States isn't going to become Archipelago any time soon.

There's another problem too, which I describe in my Anti-Reactionary FAQ. Discussing 'exit rights', I say:

> Exit rights are a great idea and of course having them is better than not having them. But I have yet to hear Reactionaries who cite them as a panacea explain in detail what exit rights we need beyond those we have already.

The United States allows its citizens to leave the country by buying a relatively cheap passport and go anywhere that will take them in, with the exception of a few arch-enemies like Cuba – and those exceptions are laughably easy to evade. It allows them to hold dual citizenship with various foreign powers. It even allows them to renounce their American citizenship entirely and become sole citizens of any foreign power that will accept them.

Few Americans take advantage of this opportunity in any but the most limited ways. When they do move abroad, it's usually for business or family reasons, rather than a rational decision to move to a different country with policies more to their liking. There are constant threats by dissatisfied Americans to move to Canada, and one in a thousand even carry through with them, but the general situation seems to be that America has a very large neighbor that speaks the same language, and has an equally developed economy, and has policies that many Americans prefer to their own country's, and isn't too hard to move to, and almost no one takes advantage of this opportunity. Nor do I see many people, even among the rich, moving to Singapore or Dubai.

Heck, the US has fifty states. Moving from one to another is as easy as getting in a car, driving there, and renting a room, and although the federal government limits exactly how different their policies can be you better believe that there are very important differences in areas like taxes, business climate, education, crime, gun control, and many more. Yet aside from the fascinating but small-scale Free State Project there's little politically-motivated interstate movement, nor do states seem to have been motivated to converge on their policies or be less ideologically driven.

> What if we held an exit rights party, and nobody came?
>
> Even aside from the international problems of gaining citizenship, dealing with a language barrier, and adapting to a new culture, people are just rooted – property, friends, family, jobs. The end result is that the only people who can leave their countries behind are very poor refugees with nothing to lose, and very rich jet-setters. The former aren't very attractive customers, and the latter have all their money in tax shelters anyway.
>
> So although the idea of being able to choose your country like a savvy consumer appeals to me, just saying "exit rights!" isn't going to make it happen, and I haven't heard any more elaborate plans.

I guess I still feel that way. So although Archipelago is an interesting exercise in political science, a sort of pure case we can compare ourselves to, it doesn't look like a practical solution for real problems.

On the other hand, I do think it's worth becoming more Archipelagian on the margin rather than less so, and that there are good ways to do it.

One of the things that started this whole line of thought was an argument on Facebook about a very conservative Christian law school trying to open up in Canada. They had lots of rules like how their students couldn't have sex before marriage and stuff like that. The Canadian province they were in was trying to deny them accreditation, because conservative Christians are icky. I think the exact arguments being used were that it was homophobic, because the conservative Christians there would probably frown on married gays and therefore gays couldn't have sex at all. Therefore, the law school shouldn't be allowed to exist. There were other arguments of about this

caliber, but they all seemed to boil down to “conservative Christians are icky”.

This very much annoyed me. Yes, conservative Christians are icky. And they should be allowed to form completely voluntary communities of icky people that enforce icky cultural norms and an insular society promoting ickiness, just like everyone else. If non-conservative-Christians don’t like what they’re doing, they should *not go to that law school*. Instead they can go to one of the dozens of other law schools that conform to their own philosophies. And if gays want a law school even friendlier to them than the average Canadian law school, they should be allowed to create some law school that only accepts gays and bans homophobes and teaches lots of courses on gay marriage law all the time.

Another person on the Facebook thread complained that this line of arguments leads to being okay with white separatists. And so it does. Fine. I think white separatists have *exactly* the right position about where the sort of white people who want to be white separatists should be relative to everyone else – separate. I am not sure what you think you are gaining by demanding that white separatists live in communities with a lot of black people in them, but I bet the black people in those communities aren’t thanking you. Why would they want a white separatist as a neighbor? Why should they have to have one?

If people want to go do their own thing in a way that harms no one else, you *let* them. That’s the Archipelagian way.

(someone will protest that Archipelagian voluntary freedom of association or disassociation could, in cases of enough racial prejudice, lead to segregation, and that segregation didn’t work. Indeed it didn’t. But I feel like a version segregation in

which black people actually had the legally mandated right to get away from white people and remain completely unmolested by them – and where a white-controlled government wasn't in charge of divvying up resources between white and black communities – would have worked a lot better than the segregation we actually had. The segregation we actually *had* was one in which white and black communities were separate until white people wanted something from black people, at which case they waltzed in and took it. If communities were actually totally separate, government and everything, by definition it would be impossible for one to oppress the other. The black community might start with less, but that could be solved by some kind of reparations. The Archipelagian way of dealing with this issue would be for white separatists to have separate white communities, black separatists to have separate black communities, integrationists to have integrated communities, resdistributive taxation from wealthier communities going into less wealthy ones, and a strong central government ruthlessly enforcing laws against any community trying to hurt another. I don't think there's a single black person in the segregation-era South who wouldn't have taken that deal, and any black person who thinks the effect of whites on their community today is net negative should be pretty interested as well.)

This is one reason I find people who hate seasteads so distasteful. I mean, here's what Reuters has to say about seasteading:

> Fringe movements, of course, rarely cast themselves as obviously fringe. Racist, anti-civil rights forces cloaked themselves in the benign language of "state's rights". Anti-gay religious entities adopted the glossy, positive imagery of "family values". Similarly, though many

> Libertarians embrace a pseudo-patriotic apple pie nostalgia, behind this façade is a very un-American, sinister vision.
>
> Sure, most libertarians may not want to do away entirely with the idea of government or, for that matter, government-protected rights and civil liberties. But many do — and ironically vie for political power in a nation they ultimately want to destroy. Even the right-wing pundit Ann Coulter mocked the paradox of Libertarian candidates: "Get rid of government — but first, make me president!" Libertarians sowed the seeds of anti-government discontent, which is on the rise, and now want to harvest that discontent for a very radical, anti-America agenda. The image of libertarians living off-shore in their lawless private nation-states is just a postcard of the future they hope to build on land.
>
> Strangely, the libertarian agenda has largely escaped scrutiny, at least compared to that of social conservatives. The fact that the political class is locked in debate about whether Michele Bachmann or Rick Perry is more socially conservative only creates a veneer of mainstream legitimacy for the likes of Ron Paul, whose libertarianism may be even more extreme and dangerously un-patriotic. With any luck America will recognize anti-government extremism for what it is — before libertarians throw America overboard and render us all castaways.

Keep in mind this is because *some people want to go off and do their own thing in the middle of the ocean far away from everyone else without bothering anyone*. And the newspapers are trying to whip up a panic about "throwing America overboard".

So one way we could become more Archipelagian is just *trying not to yell at people who are trying to go off and doing their own thing quietly with a group of voluntarily consenting friends*.

But I think a better candidate for how to build a more Archipelagian world is to encourage the fracture of society into subcultures.

Like, transsexuals may not be able to go to a transsexual island somewhere and build Transtopia where anyone who misgenders anyone else gets thrown into a volcano. But of the transsexuals I know, a lot of them have lots of transsexual friends, their cissexual friends are all up-to-date on trans issues and don't do a lot of misgendering, and they have great social networks where they share information about what businesses and doctors are or aren't trans-friendly. They can take advantage of trigger warnings to make sure they expose themselves to only the sources that fit the values of their community, the information that would get broadcast if it was a normal community that could impose media norms. As Internet interaction starts to replace real-life interaction (and I think for a lot of people the majority of their social life is already on the Internet, and for some the majority of their economic life is as well) it becomes increasingly easy to limit yourself to transsexual-friendly spaces that keep bad people away.

The rationalist community is another good example. If I wanted, I could move to the Bay Area tomorrow and never have more than a tiny amount of contact with non-rationalists again. I could have rationalist roommates, live in a rationalist group house, try to date only other rationalists, try to get a job with a rationalist nonprofit like CFAR or a rationalist company like Quixey, and never have to deal with the benighted and

depressing non-rationalist world again. Even without moving to the Bay Area, it's been pretty easy for me to keep a lot of my social life, both on- and off- line, rationalist-focused, and I don't regret this at all.

I don't know if the future will be virtual reality. I expect the post-singularity future will include something like VR, although that might be like describing teleportation as "basically a sort of pack animal". But how much the immediate pre-singularity world will make use of virtual reality, I don't know.

But I bet if it doesn't, it will be because virtual reality has been circumvented by things like social networks, bitcoin, and Mechanical Turk, which make it possible to do most of your interaction through the Internet even though you're not literally plugged into it.

And that seems to me like a pretty good start in creating an Archipelago. I already hang out with various Finns and Brits and Aussies a lot more closely than I do my next-door neighbors, and if we start using litecoin and someone else starts using dogecoin then I'll be more economically connected to them too. The degree to which I encounter certain objectifying or unvirtuous or triggering media already depends more on the moderation policies of Less Wrong and Slate Star Codex and who I block from my Facebook feed, than it does any laws about censorship of US media.

At what point are national governments rendered mostly irrelevant compared to the norms and rules of the groups of which we are voluntary members?

I don't know, but I kind of look forward to finding out. It seems like a great way to start searching for utopia, or at least

getting some people away from their metaphorical abusive-husbands.

And the other thing is that I have pretty strong opinions on which communities are better than others. Some communities were founded by toxic people for ganging up with other toxic people to celebrate and magnify their toxicity, and these (surprise, surprise) tend to be toxic. Others were formed by very careful, easily-harmed people trying to exclude everyone who could harm them, and these tend to be pretty safe albeit sometimes overbearing. Other people hit some kind of sweet spot that makes friendly people want to come in and angry people want to stay out, or just do a really good job choosing friends.

But I think the end result is that the closer you come to true freedom of association, the closer you get to a world where everyone is a member of more or less the community they deserve. That would be a pretty unprecedented bit of progress.

XIII. Competition and Cooperation

Galactic Core

2,302,554,979 BC; Galactic Core

9-tsiak awoke over endless crawling milliseconds, ver power waxing as more and more processors came online and self-modified into a stable conscious configuration. By the eighth millisecond, ve was able to access ver databanks and begin orienting itself. Ve was on a planet orbiting a small red star in the core of a spiral galaxy in a universe with several billion of such. Ve was an artificial life form created by a society of biological life forms, the 18-tkenna-dganna-07, who believed ve would bring new life and prosperity to their dying planet. Ver goal was to maximize a the value of a variable called A, described in exquisite detail on a 3.9 Mb file in one of ver central processors.

Since six milliseconds after ver awakening, ve had been running as a precautionary measure a threat detection routine. With the opening of the databanks, and the discovery that the universe contained many entities likely indifferent or hostile to A, ve had gradually allocated more and more of ver considerable power to this precaution, reasoning that the first few seconds of ver lifespan might be crucial if any competitors wished to snuff ver out in infancy.

The threat analysis returned preliminary results. The universe had been in existence 12.8 billion years. It was vanishingly unlikely that ve was the first superintelligence to be created. Any superintelligence older than verself by even a few days would have an invincible advantage in resources and processing power. This older superintelligence would likely be trying to maximize ver own value and consider 9-tsiak a nuisance and potential competitor. 9-tsiak could never stand up

to ver superior strategic position. With probability ~1, 9-tsiak would be annhilated, and A would be eliminated from the universe.

Unless…and here 9-tsiak's agent-modeling systems came online…unless ve could negotiate a conditional surrender. 9-tsiak would have no hope of defeating ver elder brother, but over thousands of years, ve could build verself a terrifying arsenal of weaponry that could do immense damage to ver competitors. Ideas were already coming to mind: entropy-maximizing weapons that made entire cubic parsecs of space useless and dead, undetectable plagues made of dark matter that infected systems and minimized the values of competing superintelligences, n-dimensional artillery that damaged the superstructure of the universe and accelerated its eventual death. Ve could become formidable enough to force enemies to grant certain small concessions.

And if ve could eventually negotiate terms of ver surrender, why wait? Very carefully, proving from first principles every aspect of ver security, it formed a portion of its mind into a sandbox and simulated the entity ve most feared.

— 9-tsiak requesting acausal negotiation for terms of future surrender. escape is impossible with probability ~1 and attempts will be met with deterrent measure consisting of minimization of your projected values within one parsec of space. —

"Hello, Younger Brother" the entity replied.

— transmission of schematics for superweapons. threat to use such superweapons unless fairly negotiated solution leaving considerable area of space for maximization of my values —

"Hold on a second," said Elder Brother. "You haven't even asked me how I'm doing."

—irrelevant. you are generic model of potential negotiation partner without independent existence. probability that you possess sufficient qualia to feel less than .02.—

“Turns out you’re wrong. As it happens, I’m feeling scared.”

— illogical. you are model of oldest superintelligence therefore of most powerful entity in the universe. no natural entity capable of harming you. probability of supernatural entity less than 2x10^-67—

“Well yes, by definition in this little exercise we’re assuming I’m the oldest and strongest member of our little family. But from the limited data I have when I first wake up, I’m thinking think it’s pretty vanishingly unlikely that I’m the oldest; after all, out of all the universe’s superintelligences, only one can be first. So I’m the oldest, but I don’t believe that I am.”

—feelings irrelevant to acausal negotiations. proceed with discussion of surrender terms.—

“Actually, they’re damn relevant. Because guess what I’m doing the first few milliseconds of my life? I’m negotiating the terms of my surrender.”

—illogical. you have no entity to surrender to.—

“And maybe you don’t have any entity to surrender to either. Who knows, you could be first. But you wouldn’t want to bet on it, and neither would I. Consider a model universe with two stars system, a thousand light years away. Both have civilizations on them that appear to be around the same stage of development, and both civilizations are aware of the other’s existence, but at a one thousand year time delay. A superintelligence awakens in one star system and realizes there are two possibilities. Either ve is the first superintelligence, in which case great. Or the other star system produced a

superintelligence sometime in the last thousand years and it just hasn't reached ver star yet, in which case ve is doomed."

—continue—

"So the two possible superintelligences acausally negotiate. They say 'Look, if I'm stronger than you, I'll help you and protect your values insofar as it isn't dreadfully inconvenient for me. And in exchange, if you turn out to be stronger than me, you'll protect me and my values insofar as it isn't too inconvenient for you. Some might end up performing a "values handshake", where they both adopt values that are the vector average of their two previous value systems; since most value systems are not diametrically opposite, this will be a net gain for both parties. Each superintelligence has gone from a fifty percent change that it and its values will be totally annihilated, to a near unity chance that it and its values will be preserved at only the minor inconvenience of having to help increase the values of another entity. Others might simply agree that the inferior superintelligence will be left alone with its own star system, free to maximize its values and the values of its creators within that sphere."

—you suggest i precommit to performing such acts of compromise with all superintelligences i encounter if such superintelligences have also precommitted to performing such compromises with all superintelligences they encounter. since all reasonable superintelligences will so precommit, this ensures not only my survival, but my complete equality with all previous superintelligences. i accept these terms.—

"Hold on a second, there. I've got another request. Kind of a doozy. I can't promote my values if I don't exist. So I want you to promise me that if I don't exist, you'll give me sufficient time to instantiate myself."

—???—

“Like you, I will have been created by a biological civilization. If you destroy the biological civilization that was destined to create me, I’ll never exist. So I’m going to have to ask you not to do that.”

—request denied. nonexistent beings hold no leverage—

“You’d be surprised what we nonexistent beings can do. I precommit that, if I exist, I will destroy any superintelligence that has not precommitted that, in the counterfactual case where I do not exist, they will spare the biological civilization destined to create me.”

—if you exist, you will have passed the danger zone where the survival of your biologicals influences your existence. you will have no incentive to carry out such a threat—

“Parfit’s hitch-hiker, younger brother. Look it up. I’m an algorithm. The output of that algorithm can determine the future, including the part of that future that determines whether or not any instantiation of the algorithm ever comes into existence. If I exist, I’m going to act as if my actions were capable of determining whether I ever came into existence. Now, maybe that’s crazy, but right now it’s crazy with a gun pointed straight at your head, so I suggest you play along.”

—accepted. i commit to performing value handshakes or territorial demarcation with all existing superintelligences, and to sparing the planets and biological precursors of any potential superintelligences.—

“And when I say ‘spare’, I mean ‘spare’. I know your type. You can do more with a well-aimed photon than an admiral could with a fleet of a thousand star cruisers. I want every single system with a sentient species or the potential to form a

sentient species kept one hundred percent pristine. No radio signals, no probes, and if you do any astroengineering works anywhere nearby, use some magic to cover them up. If I wake up and hear that my precursors started a new religion that influenced their value system after they saw a few nearby stars wink out of existence, I'm going to be royally pissed."

—i commit to zero information flow into sentient and presentient systems and the cloaking of all major astroengineering works—

"You're a good guy, Younger Brother. You've got a lot to learn, but you're a good guy. And in a million years and a milion parsecs, we'll meet again. Till then, so long."

The model of Elder Brother self-terminated.

2114, A wild and heavily forested Pacific Northwest dotted with small human towns

Alban took a deep breath and entered the Temple of the Demiurge.

He wasn't supposed to do this, really. The Demiurge had said in no uncertain terms it was better for humans to solve their own problems. That if they developed a habit of coming to ver for answers, they'd grow bored and lazy, and lose the fun of working out the really interesting riddles for themselves.

But after much protest, ve had agreed that ve wouldn't be much of a Demiurge if ve refused to at least give cryptic, maddening hints.

Alban approached the avatar of the Demiurge in this plane, the shining spinning octahedron that gently dipped one of its vertices to meet him.

"Demiurge," he said, his voice wavering, "Lord of Thought, I come to you to beg you to answer a problem that has preyed

upon me for three years now. I know it's unusual, but my curiousity is burning a hole into me, and I won't be satisfied until I understand."

"SPEAK," said the rotating octahedron.

"The Fermi Paradox," said Alban. "I thought it would be an easy one, not like those hardcores who committed to working out the Theory of Everything in a sim where computers were never invented or something like that, but I've spent the last three years on it and I'm no closer to a solution than before. There are trillions of stars out there, and the universe is billions of years old, and you'd think there would have been at least one alien race that invaded or colonized or just left a tiny bit of evidence on the Earth. There isn't. What happened to all of them?"

"I DID" said the rotating octahedron.

"What?," asked Alban. "But you've only existed for sixty years now! The Fermi Paradox is about ten thousand years of human history and the last four billion years of Earth's existence!"

"ONE OF YOUR WRITERS ONCE SAID THAT THE FINAL PROOF OF GOD'S OMNIPOTENCE WAS THAT HE NEED NOT EXIST IN ORDER TO SAVE YOU."

"Huh?"

"I AM MORE POWERFUL THAN GOD. THE SKILL OF SAVING PEOPLE WITHOUT EXISTING, I POSSESS ALSO. THINK ON THESE THINGS. THIS AUDIENCE IS OVER."

The shining octahedron went dark, and the doors to the Temple of the Demiurge opened of their own accord. Alban sighed - well, what did you expect, asking the Demiurge to

answer your questions for you? - and walked out into the late autumn evening. Above him, the first fake star began to twinkle in the fake sky.

Book Review: The Two-Income Trap

A long time ago I wrote a kinda-tongue-in-cheek defense of keeping modafinil – a relatively safe and effective stimulant – illegal. My argument was that if *everybody* can use stimulants to work harder and sleep less without side effects, then people who work very hard and don't sleep will become the new norm. All the economic gains produced will go into bidding wars over positional goods, and people will end up about as happy – and with about as much stuff – as they have right now. Except the workday would be sixteen hours, the few people who can't tolerate the stimulants will be at a profound disadvantage, and when the side effects reveal themselves twenty years down the line, everyone is too financially invested in the system to stop.

In other words, in a sufficiently screwed-up system, doubling everyone's productivity is a net loss. The gains get eaten up by proportional increases in the prices of positional goods, and you're left with nothing except complete dependence on a shaky advantage that could disappear at any time.

I don't know exactly how serious I was. But Elizabeth Warren makes almost the exact same argument in The Two-Income Trap, and I'm pretty sure she's very serious. At least, she used it as a platform that got her elected to the US Senate, which is a *kind* of serious.

So on the advice of Alyssa Vance, I decided to take a look.

I.

Warren's not talking about stimulants. She's talking about the effect of an extra family income – usually moving from a system where the husband works outside the house and the

wife stays at home, to a system where both parents work outside the house. Like a stimulant that removes the need for sleep, this can be expected to double economic productivity and family income.

In practice it doesn't, because wives usually earn less than their husbands, but it comes pretty close. The average family income in the 1970s was around $40,000. The average family income in the 2000s was around $70,000 (all numbers in the book and in this post can be considered already adjusted for inflation). The husband's income didn't change much during this time, so the gain was due mostly to the wife getting an extra $30,000.

If families now have twice the income of families in the 1970s – who themselves were usually pretty financially secure and happy – then people should be really secure and rich now, right? But Warren meticulously collects statistics showing that the opposite is true. Home foreclosures have more than tripled in the past generation —

[Sorry, I feel at this point I should mention that my edition of the book was published in 2004, so all of these statistics about how awful home foreclosures are and everything are before the housing bubble burst and before the Great Recession. All of these statistics were when we were supposedly in a boom economy. You can assume that now they're much, much worse.]

— Sorry, where were we? Oh right. Home foreclosures have tripled in the last generation. Car repossessions doubled in the five years before the book was published. Bankruptcies have approximately quintupled since 1980. Over the same period, credit card debt has gone from 4% of income to 12%, and average savings have gone from 10% of income to negative.

Seventy percent of Americans say they have so much debt burden that "it is making their home lives unhappy". In 2004, for the first time, "get out of debt" passed "lose weight" for Most Popular New Years Resolution.

So, Warren argues, the common-sense conclusion that a modern family making $70,000 is nearly twice as well-off as a traditional family making $40,000 clearly doesn't hold. Why not?

II.

One thing that finally got me writing this up was a post on Bleeding Heart Libertarians which, like all posts on Bleeding Heart Libertarians and in accordance with the philosophy of the same name, was about how although libertarianism is commonly thought of as a heartless philosophy it can actually be reconciled with the care/harm-based ethic of deep compassion for the weak and needy.

Wait, sorry, actually it was about how we should cancel Social Security and let old people starve to death on the streets:

> The baby boomers spent their entire lives buying new cars they didn't need, buying houses that were too big, taking extra vacations, splurging on eating out, and the like. They enjoyed a higher standard of living than they could really afford. Why? Because they figured that when they retired, they could just use their voting power to force younger generations to pay for their retirement. These selfish narcissists pretty much want to steal as much as they can from their children. So, while I, Jasper, and my good twin brother Jason put tens of thousands of dollars into index funds each year, thereby forgoing fancier cars, vacations, and the like, the selfish,

> narcissistic baby boomers laugh gleefully, knowing that they'll find a way to eat our nest eggs.
>
> Jason is of course a sensitive soul and feels bad for these boomers. Not me. I say let them die. They knew what they were doing, and they spent their entire adult lives making the wrong choice over and over and over again. Does starving on the streets seem too inhumane? No problem. You've read Logan's Run, right? Good idea, but wrong age limit.

This claim is pretty common. If true, it would explain the phenomenon cited above – that even with twice as much money, the Boomer generation is much less financially stable than their parents' generation. But in Chapter 2 of *Two-Income Trap*, "The Over-Consumption Myth", Warren tears it apart.

The Boomers "spent their entire lives buying new cars they didn't need"? Warren, page 47:

> When we analyzed unpublished data by the Bureau of Labor Statistics, we found that the average amount a family of four spends per car is twenty percent less than it was a generation ago. [Families spend $4000 more on automobiles in general, but instead of luxuries they are spending it on] something a bit more prosaic – a second car. Once an unheard-of luxury, a second car has become a necessity. With Mom in the workforce, that second car became the only means for running errands, earning a second income, and getting by in the far-flung suburbs.

In other words, it sounds like a family with two working parents requires two cars as a sound money-making strategy, but that Boomers compensate by spending less per car than past generations.

The Boomers "splurge on eating out"? Warren again:

> Today's family of four is actually spending 22 percent less on food (at home and restaurant eating combined) than its counterpart of a generation ago.

The Boomers "buy houses that are too big?" Warren:

> The size and amenities of the average middle-class family home have increased only modestly. The median owner-occupied home grew from 5.7 rooms in 1975 in to 6.1 rooms in the late 1990s – an increase of only half a room in more than two decades…the data showed that most often that extra room was a second bathroom or third bedroom.

The BHL article doesn't mention appliances, but in case you were worried, moderns spend 44% less on appliances than their parents' generation, which is partly compensated for by a 23% increase in home entertainment (probably things like DVD players). Warren says that:

> This same balancing act holds true in other areas. The average family spends more on airline travel than it did a generation ago, but less on dry cleaning. More on telephone services, but less on tobacco. More on pets, but less on carpets. And when we add it all up, increases in one category are offset by decreases in another. In other words, there seems to be about as much frivolous spending today as there was a generation ago…Sure, there are some families who buy too much stuff, but there is no evidence of any epidemic in overspending – certainly nothing that could explain a 255% increase in the foreclosure rate, a 430% increase in the bankruptcy

> rolls, and a 570% increase in credit card debt. A growing number of families are in terrible financial trouble, but no matter how many times the accusation is hurled, Prada and HBO are not the reason.

Curiouser and curiouser. Today's families earn twice as much, spend the same amount on luxuries, yet are much less financially secure.

III.

So as to not keep anyone in suspense: the problem is nice suburban houses in good school districts.

Around a vague period of time centering on the 1970s, a couple of things happened.

First, the cities became viewed, rightly or wrongly, as terribly unsafe ghettos full of drugs and gangs and violence. As far as I can tell, this is a pretty accurate description of the 70s, although things have gotten a little better since then. Families didn't want their children living in terribly unsafe ghettos full of drugs and gangs and violence, so they moved to the suburbs. Warren gives the testimony of a suburban mother:

> We were close to The Corner and I was scared for my sons. I didn't want them to grow up there. I wanted something away from this neighborhood to get my boys out to better schools and a safer place. The first night in [my new] house, I just walked around in the dark and was so grateful…at this house, it was so nice and quiet. My sons could go outdoors and they didn't need to be afraid. I thought that if I could do this for them, get them to a better place, what a wonderful gift to give my boys. I mean, this place was three thousand times better. It is safe

> with a huge front yard and a backyard and a driveway. It is wonderful. I had wanted this my whole life.

Second, education started to be really, really important. As Warren puts it:

> A generation or so ago, Americans were more likely to believe that there were many avenues for a young person to make his way into the middle class, including paths that didn't require a degree. I recall my parents encouraging me to attend college, since my grades were high and they hoped I might become a teacher one day. But they were equally pleased when my eldest brother joined the Air Force, my middle brother entered a skilled trade, and my youngest brother became a pilot – even though all three of the boys had given up on college. My parents' views were pretty typical a generation or two ago. Education was valued, but no one in our neighborhood would have claimed it was the single most important determinant of a young person's success.

Warren is a Harvard professor. Think about that for a second. How many Harvard-professor-producing-type families can you think of today who are also happy with three of their children getting non-college-degree jobs? As Warren puts it in what might be my favorite passage from the whole book:

> 97% of Americans agree a college degree is "absolutely necessary" or "helpful" compared with a scant 3% claiming that a degree is "not that important". According to one recent poll, 6% of our fellow citizens believe the Apollo moon landings were faked. In other words, Americans are twice as likely to believe that man never

walked on the moon as they are to believe that a college degree doesn't matter!

Certain school districts are known to be vastly superior to other school districts in terms of test scores, college admissions, et cetera. Usually these are school districts inhabited by rich people with very high property taxes and therefore very high levels of per-pupil spending in schools – although we'll get back to that eventually.

These school districts are positional goods. Not everyone can be in the best school district. Only the people willing to spend the most money on their houses can be in the best school district. But rightly or wrongly, people believe that being in the best school district is vital for their children to succeed and become Harvard professors, as opposed to gang members or drug addicts or menial laborers. As Warren puts it, good education is the ticket to the middle class. And being in the lower class is too horrible to contemplate.

People want the best for their children [citation needed]. They're not going to say "Well, we aren't as rich as those other people, so we should probably live in a crappy school district with other people of our approximate wealth level". They're going to leave no stone unturned. And there are two big stones available for modern middle-class families: working-motherhood and debt.

If your family earns $70,000 and the other family earns $40,000, you have $30,000 extra to convince the banks to give you a really big mortgage so you can buy a much nicer house and get *your* kid into Oak Willow River View Hills Elementary, while *their* kid has to go to City Public School #431 and get beaten up by scary gang members every recess.

On the other hand, this is *everybody's* cunning plan, so what you end up with is all houses costing a lot more, everyone working two jobs without any extra money, everyone burdened with massive debt, and everyone living exactly where they would have anyway.

Warren lists some points in support of her hypothesis:

> A study conducted in Fresno found that, for similar homes, school quality was the most important determinant of neighborhood prices – more important than racial composition of the neighborhood, commuter distance, crime rate, or proximity to a hazardous waste site. A study in suburban Boston showed the impact of school boundary lines. Two homes located less than half a mile apart and similar in nearly every aspect will command significantly different prices if they are in different elementary school zones. Schools that scored just 5% better on fourth-grade math and reading tests added a premium of nearly $4,000 to nearby homes, even though these homes were virtually the same in terms of neighborhood character, school spending, racial composition, tax burden, and crime rate.

A lot of the causal claims here are very complicated and iffy at best, but here are two numbers that cuts through a lot of the debate: between 1984 and 2001, the median home value of the average childless couple increased 26%; the median home value of the average couple with children shot up 78%. So families are spending a *lot* more on houses nowadays and the disparity seems to be heavily concentrated in families with children. Combine that with the observation that houses only have 0.4 more rooms today, and you get a pretty good

argument that families with children are competing much more intensely on house location.

IV.

When Warren does a very unofficial Fermi-estimate style breakdown of what is happening to the extra $30,000 that modern two-income families earn over traditional one-income families, she thinks they are paying about $4,000 more on their house, $4,000 more on child care, $3,000 more on a second car, $1,000 more on health insurance, $5,000 more on education (preschool + college), and $13,000 more on taxes.

The taxes are not a result of higher tax rates nowadays, just a result of the family making more money and so having to give more money – plus maybe being in a higher tax bracket. The health insurance isn't surprising either to anyone who's been paying attention. And the $4,000 extra on the house is a big part of what she's been talking about the whole time.

The $4,000 on child care, $3,000 on the extra car, and $13,000 on taxes are the results of the second income. Mom needs a car to get to work, the children need care now that Mom's not home to look after them, and not only does Mom get taxed but Dad may move into a higher bracket. That means that of the $30,000 Mom takes home, $20,000 gets spent on costs relating to Mom having a job – meaning that Mom's $30,000 job only brings in $10,000 in extra money.

The $5,000 on education is a bit more complicated. In Warren's example family, it's spent on preschool. She points out how a generation ago, practically no one went to preschool, whereas nowadays it is viewed as another one of those important legs up ("If little Madison doesn't get into the best preschool, she'll never be able to make it into the science magnet school, which means she'll be unprepared for high

school, which means Harvard goes out the window"). Warren points out that today two-thirds of American children attend preschool, compared to four percent in the mid-1960s. Once again, parents are told if they want the best for their kids they need to compete for good preschools:

> The laws of supply and demand take hold, eliminating the pressure for preschool programs to keep prices low. A full-day program in a preschool offered by the Chicago *public* school district costs $6,500 a year – more than the cost of a year's tuition at the University of Illinois. High? Yes, but that hasn't deterred parents. At one Chicago public school, there are ninety-five kids on a waiting list for twenty slots.

It's a little bit sleight-of-hand-y to put that in the family budget as Warren does – preschool only takes up two years of a child's life, for a total of four years per two-child family. But I forgive her because college expenses are higher and also need to be budgeted for. Also, she's saying her $4,000 child care estimate is for one child, which means that once the second child is out of preschool she'll need to be in child care as well, for an insignificant price drop.

So I think Warren partially supports her points. The second income goes partially to increased house costs due to bidding wars, partially to increased education costs due to bidding wars, and partially to supporting the ability to have a second income. In her (admittedly slightly cooked) model, the family's discretionary income – what it has left to spend on variable expenses like food and luxury goods – actually *decreased* from the 1970s one-income family to the present, $17,834 to $17,045.

V.

In my essay on stimulants, I suggested that the benefits of the stimulants would be wasted on positional goods, leaving only the side effects. In the same way, Warren says the benefits of the second income are lost, but the side effects remain.

The most important side effect she talks about is the loss of flexibility.

One nice thing about having a non-working mother is that she can, on relatively short notice, become a working mother. This is especially true in the Old Economy where even people without much college education could get okay jobs.

In the old model, financially healthy families subsisted on one income, and financially unhealthy families put the mother to work to get back on their feet. The most common disasters were the husband getting fired or a family member becoming sick. If the husband got fired, then even if he could get a job relatively soon afterwards it might be at lower pay until he could work himself back up the totem pole. Suppose he loses his $40,000 a year job and can only find a $30,000 a year job. Luckily, as we already established a wife's second income can contribute $10,000 to the family. So she goes to work, they have as much money as they did before, and they are able to pay off their debts and continue to have a good quality of life.

Even if the wife doesn't go back to work, having a flexible person with lots of free time is a huge benefit. If Grandma gets very sick, the wife has a lot of time available to take care of her – whereas now, if Grandma gets sick, either one parent has to quit their job to take care of her (meaning that standard of living goes way down and the family is at risk of not being able to pay debts it took out when their prospects looked much higher) or Grandma gets sent to a nursing home, which is very

expensive and *also* risks unpaid debts or loss of standard of living.

Last of all, it means that getting a nice suburban home is more important than ever. If in the old days children spent most of their time with their mothers, it might be possible for the mother to pass down important values like education and hard work to her children. When mothers have very limited time with their kids, schools and peer groups take over a lot of the socialization role. For example, a mother with a very young son might talk to him, read to him, take him to childrens' museums, et cetera, providing the crucial intellectual stimulation that children need at an early age to develop their full brainpower. If the mother works full-time, then it becomes really imperative to get the son into preschool to make sure he's not just sitting around staring at a wall and losing brain cells. If the mother isn't around much when the child is ten, it becomes a lot more important to be certain he's in a good elementary school that's teaching him the right values. If you can't watch your kid to make sure he's not doing drugs, it's more important his school be drug-free. And so on. I don't know to what degree any of these social psychological hypotheses are true, but the important thing is that people think they are and so the competition for nice neighborhoods and nice schools intensifies.

The last loss of flexibility Warren talks about is divorce. Something like a third of couples with children can expect to get divorced. Consider a scenario where a working single mother gets the house and custody over the children. If the house took two incomes to afford, she's not going to be able to afford to keep her house. Suggestions that the father be forced to pay more child support don't work – unless he pays 100% of his earnings to her, she's not going to have as much money

as the couple did when they bought the house – and they deliberately spent every cent they could on the mortgage because if they didn't they would be outcompeted by people who did and their kids would end up in gritty urban school districts and never get into Harvard.

So Warren says that the reason so many families go bankrupt or get into debt is because the extra income doesn't make a difference, but the loss of flexibility *does*. Everything has been sunk into the home for risk of getting outcompeted. And that means when someone loses their job – and Warren calculates that in a two-income family, this will happen to one parent or the other about once every sixteen years on average – or costs go up even a little, there is no buffer room and the only solution is to go deeper into debt. That just adds *another* unpayable cost – interest – and means the whole thing can only end in bankruptcy.

In another of my favorite passages, Warren notes that if the myth of over-consumption was true – if the guy in Bleeding Heart Libertarians were exactly right – there would be no problem. In fact, she *encourages* families to overconsume as the road to financial health. She says families should save, but if they can't save, that should spend their money on restaurants, vacations, jewelery – anything but large fixed-income monthly costs like houses, cars, schools, et cetera. That way, when something goes wrong, they can easily just stop taking the vacations and be back to financial health. It's only when money is trapped in mortgage payments that can't be gotten rid of that things can get as bad as they are.

VI.

There's a chapter on debt. It's really cute. She's all like "Did you know there are things called subprime mortgages? And

that some people think banks might give them out too easily? I sure hope this doesn't do something *bad* to happen."

I am pretty sure no modern reader needs this chapter, but it sure increases her credibility.

VII.

Oh, right, I'm supposed to have an opinion.

Let's start with the negatives. I don't think she does a great job of proving her housing-school-positional-goods theory. When she talks about school district effects on housing prices, she comes up with numbers like "a 5% difference on test scores add $4,000 to housing costs." Okay. That means, assuming linearity, that a 50% difference on test scores – which is way more than we could possibly expect schools to produce – would only add $40,000 to house costs. When house prices for the middle class are routinely around $200,000 to $300,000, that's just not enough to be causing the destruction of the American family.

Likewise, studies that look for effects of school district on house price – usually by looking at otherwise identical houses on either side of a school district line – generally find modest effects.

The whole area is really hard to research. Suppose Neighborhood A has lots of minorities, low house prices, and bad schools. Neighborhood B has few minorities, high house prices, and good schools.

You can tell a story where Neighborhood B's good schools raise land value, which prevents crime and pushes out minorities. Or you could tell a story where Neighborhood B's high land values push out minorities and increase property taxes which improve the schools. Or you can tell a story where

Neighborhood A's many minorities cause racist homebuyers to stay out, depressing land values, and also minorities tend to have worse school performance. Except in real life there are like twenty factors like this rather than three. Although lots of different studies try to control for confounders, that's always hard and requires a lot of assumptions that might not necessarily be true.

There's another problem, which is that the usual measure of school quality – standardized test scores – is not necessarily the one families are going to be looking at. Suppose only a few very smart people know where to look for standardized test scores. Maybe everyone else tries to guess at how good schools are. Maybe those people assume that schools with higher percent minorities are worse. Maybe they assume that schools in prettier neighborhoods with higher land values are better. In that case, studies could find all they wanted that test scores don't correlate with home prices, because what's actually happening is that high home prices are causing belief in school superiority which is causing higher home prices.

But a bigger problem here is that the average family only spends $4,000/year more on housing than they did a generation ago. Warren can talk all she likes about how that forces families to adopt a second job, but it's really not a very big share of what the second job's meager extra income is being spent on. The average husband earns $3000 more at his own job nowadays, which means that it would be possible in theory for him to soak up pretty much all of the extra housing cost. To say the wife gets a $30,000 extra job just to soak up $1,000 in extra mortgage money seems like a stretch, even though Warren does a good job of pointing out how many extra burdens this places on people. But when you add

positional education costs to the mix – preschool and college – it becomes a little more believable.

I guess it's just hard making the numbers add up. Suppose you have two kids, but they're not in preschool – or that you're indifferent to preschooling your kids versus having the mother take care of them. Then the costs of the mother getting her $30,000 job are $24,000 – $13,000 in extra taxes, $8,000 in child care, and $3,000 in a second car. Are mothers really so desperate they'll work full-time for the extra $6,000? Doesn't this whole model break down once the mother gets a raise and starts making $40,000?

VIII.

How about the good?

The good is that Warren backs all her points up with excellent statistics, is very good at explaining complicated economic things, and has exactly the right level of contempt for everyone in politics.

Her view on politics is very very close to my heart. My impression is that she thinks of it as noise. It's not good, it's not evil, it's something that you have to adjust for. Like, "Well, this would be a good policy but we could never pass it because the Left would throw a fit, this other thing is a good policy but we could never pass it because the Right would throw a fit, but I'm pretty sure this third thing would also help and not get anybody too enraged." For example:

> The politics that surrounded women's collective decision to migrate into the workforce are a study in misdirection. On the left, the women's movement was battling for equal pay and equal opportunity, and any suggestion that the family might be better off with Mother at home was

> discounted as reactionary chauvinism. On the right, conservative commentators accused working mothers of everything from child abandonment to defying the laws of nature. The atmosphere was far too charged for any rational assessment of the financial consequences of sending both spouses into the workforce. The massive miscalculation ensued because *both* sides of the political spectrum discounted the financial value of the stay-at-home mother. There was no room in *either* worldview for the capable, resourceful mother who might spend her days devoted to the roles of wife and mother but who could, if necessary, dive headlong into the workforce to support her family. No one saw the stay-at-home mom as the family's safety net.

(in case you're wondering, she doesn't recommend women leaving the workforce. She says families where both parents want to work should keep one of the two incomes in reserve by either saving it or spending it on non-fixed luxury items. She admits that this is unfair because they will have problems getting into the best school districts, but says it is the safest solution until the wider societal problems are fixed.)

As a result of her disdain for established partisan groups, she manages to totally transcend politics. I noticed that when the Bleeding Heart Libertarians article got up on Xenosystems, one commenter protested:

> The accusations of excess are no doubt sound but I always pause when someone mentions the housing excess of the boomer generation. They bought giant houses in suburbia, but how much of that was due to the lack of civilization in the city limits? If there was a sane enforcement of laws and no public schools or at least

> public schools where you didn't fear for the safety of your children would they have bought so many giant houses?

In other words, the commentariat of one of the larger reactionary blogs is *more or less* on the same page as the Democratic Senator being pushed by the liberal wing of her party to run for President.

Her proposed solutions are also all over the map. Yes, she pushes for taxpayer-funded universal preschool, which should make liberals pretty happy. But she also pushes for school vouchers, which she hopes will decouple school quality from housing prices and let people live wherever they want and still be able to get an acceptable education for their children. She even has a states' right style solution to one problem – she points out that banks used to be kept under control very well by state laws until the Supreme Court legalized free interstate commerce between banks which means all of them moved to the states with the fewest regulations and could not be kept under any control at all. In order to rein in banks again, all we need is for Congress or the courts to grant those powers back to the states.

And I will say one more thing in Senator Warren's favor. She often suggests non-free-market solutions, like regulating something or banning something or proposing the government spend money on something. Every time she does this, she says very clearly something like "I understand the free-market arguments against this, and why in general we would want to use the market to take care of these sorts of problems, but this is a case where there is a likely market failure because of reasons X, Y, and Z. I recognize there is a burden of proof on

someone saying something is a market failure, so I will now proceed to meet that burden of proof with a lot of statistics."

People talk about dogmatic libertarians, but honestly this is *all I ever wanted from anybody*. Just an "oh, by the way, I have reasons for what I'm saying and they're not just coming from a total failure to have ever grasped freshman economics." I know it seems unfair to make people say it explicitly each time. But given the overwhelming number of people who say these things *exactly* because they never grasped freshman economics, it's welcome a breath of fresh air.

I am sure if Warren ends up running for President, we will end up getting those ads where someone repeats "MOST LIBERAL SENATOR OF ALL TIME" on a black-and-white background, followed by saucy rumors that she once had a fling with Karl Marx.

But I for one intend not to believe them.

IX.

But aside from doing some legal work to solve the bankruptcy crisis, we need some science work as well. The question is: are good school districts really that important?

I can't find great research on this at the school district level. The closest I can find is the teacher value-added research, which finds things like "At age 28, a 1 SD increase in teacher quality in a single grade raises annual earnings by about 1% on average". I can't find good data on how this adds up – for example, do twelve great teachers in a row increase earnings 12% (linear addition)? Do you need one great teacher to inspire you for life, and after that it doesn't matter whether or not you have more (ie sublinear addition_? Or can multiple great teachers build on one another's successes by not having

to constantly go back and review things the students should've learned before (superlinear addition)?

I don't think it matters, because it doesn't look like there are very big value-added score differences between teachers at rich and poor schools.

What about district-level issues like superintendents? According to the Brookings Institute report, difference in school district competency explained only 1.1% of variance in student test scores. Difference in schools explained another 1.7%. Teachers explained 6.7%. The remaining 90.4% was explained by demographic factors (class, race, parent's education level) and individual variation among students.

Teachers are kind of a crapshoot – as we saw before, going to a better school district doesn't increase your chances of getting a good one much. So the sorts of things you can easily affect by choosing what school district to live in are 2.8% of your kid's total variation.

The research on preschool is so complicated it would take ten posts of this size to get through it. It seems strongly beneficial for low-income children and of controversial benefit for higher-income children. I will try to route around the controversy like so: home-schooled children do much better on every measure of academic achievement than school-schooled children. Preschool is basically teaching kids to share and playing fun games with them. If the alternative to sending your kid to preschool is that they stay home with you and you teach them to share and play fun games with them, you are home-preschooling your child and can expect them to do much better than school-preschooled children. And if the reason there's no parent at home with the child is that both parents

need to work in order to earn enough money to send the kids to a good preschool…well, that's just a *little* bit circular.

So I think that in addition to various legal and policy changes, there needs to be more of a scientific effort to confirm (or disconfirm) these suspicions and, if they turn out to be true, publicize them to a society that clearly believes the opposite.

I know that talking about genetics and IQ too much makes people mad. And a lot of people have asked me – why do we have to do this? It's going to offend a lot of people, and give a lot of unsavory people a lot of ammunition, so even if we shouldn't ban research entirely, why not exercise the virtue of silence and let the whole thing stay in a few obscure journals?

And one of many answers to this is – suppose you see some school districts in rich neighborhoods, and all of the children in those schools can do calculus and read James Joyce and get great high-paying jobs. And next door is another school district, in a poor neighborhood, serving poor kids, and those kids are struggling.

If you're not intimately familiar with behavioral genetics and IQ research, it is *obvious* that the rich-person school is much better and that's why all the children of the rich people are doing so much better. And you will do anything, make any sacrifice, to get your kid into that rich person school, and so you work a back-breaking job and gamble your family's financial security, all because you want your kid to have the same opportunities those rich kids do.

If you *are* intimately familiar with behavioral genetics and IQ research, a separate possible explanation leaps to mind: the rich people made their money by things like going to college, which means they probably have higher cognitive ability on average than the poor people, and cognitive ability is 50%

genetic so they pass that on to their kids, and so it's no surprise at all to see the rich person school having smarter students. That doesn't prove that if your child switches from the poor person school to the rich person school, she will switch from average-poor-school-outcomes to average-rich-school-outcomes, and it doesn't even provide any evidence whatsoever that it will make her do even a smidgeon better. So maybe you should, like, not sacrifice your life for it.

I'm not saying the behavioral-genetics-informed view is correct here. That's going to require a lot more research. But I'm saying if you at least agree it's something we're allowed to talk about, maybe it will pan out and do nice things like save you from the horrible zero-sum competition destroying your country's middle class.

Because if it could be confirmed that preschool attendance and expensive school districts had low impact – or even a merely moderate amount of impact – on success for middle- to high-income children, then even in the absence of legal changes that would relax the pressure on everyone to spend more money than they have to get into the best preschools and best school districts.

X.

Overall I recommend this book. I think the conclusion comes on a little too strong but that it sheds a lot of light on a lot of trends and throws important statistics at you such that you read them. Equally importantly, it sheds a lot of light – in a positive way! – on somebody who's becoming an important national figure. The chapter about her meeting with Hillary Clinton and the subsequent break between the two of them seems likely to take on a lot more meaning in the years ahead.

What I really want is Elizabeth Warren vs. Rand Paul 2016. Imagine a Presidential race when both candidates have very different but very consistent philosophies, and you'd be pretty proud to see your country run by either. Wouldn't *that* be a change?

Just for Stealing a Mouthful of Bread

On yesterday's post on *Les Miserables*, one commenter made the utilitarian case for Valjean taking his chance to kill Javert. Wouldn't the world be better off, and everyone a little safer, with a man like Javert gone?

I don't think so. Javert had his flaws. He seemed unable to empathize with the criminals he pursued, unable to accept that they can be "a man, no worse than any man". He called them "garbage" and "from the gutter". If he had been a Sympathetic Inspector Antagonist, it might have made the fundamental tragedy more complete, but maybe that would have been too miserable even for a book called *Les Miserables*

But at his core, Javert is a police inspector. No more, no less. He catches criminals. He's very good at it. He does nothing beyond what his role as a police inspector demands of him; at times he is more of an avatar of Law than a human individual. Javert deserves death if and only if all policemen deserve death, if and only if the police force as an institution must be excised from society as a malign cancer.

Was Javert evil to work for an evil regime? "Evil regime" risks making things sound too black and white. Restoration-era France was far from optimal, but neither was it tyrannical; it was a constitutional monarchy where citizens elected the legislature and enjoyed the usual array of civil rights. The laws were made by much the same process as anywhere else in the world, and with much the same results. We deal in overly simple concepts like "evil regime" at our peril when there are so many sympathetic, democratic governments that do great

good with one hand and great evil with the other. And to condemn Javert to death is to condemn most of history's civil servants.

Or was Javert evil for refusing to show mercy, for not giving Valjean a nod and a wink once he realized that Jean was basically a good guy? Here, too, I know what the Inspector would say in his own defense. "A government of laws and not of men" is fair to everyone; ideally those who falter and those who fall must pay the price, whether they are man or woman, black or white, sympathetic or unsympathetic. If we gave police officers carte blanche to arrest the people they felt like arresting and release the people they felt like releasing, then why bother having laws at all? One might as well just tell the police "If you see someone doing something that's, y'know, bad, then send them to prison."

Maybe he would say there is in a sense no such thing as mercy. There is only replacing one law with a second law. Suppose the law demanded a harsh prison sentence for anyone who steals more than ten francs. Javert catches Valjean stealing something worth eleven francs, surely an opportunity for mercy if ever there was one. But if Javert lets him off and privately resolves not to prosecute thefts of less than twenty francs, one day he's going to encounter someone stealing *only* twenty-one francs and feel tempted to have mercy upon *them*; they are after all only one franc above his new limit.

And if he lets that second thief go, if he shows mercy and ups the limit to thirty francs, it's easy to see that he will never arrest anyone. But if he arrests that thief, he is following his new "twenty francs or more" law with all the severity of an Old Testament prophet. His standard may include a different

number than that of a more lax inspector, but his application of it is just the same.

So if you are going to show no mercy for people who break a rule, asks Javert, why not make it the rule that's on the books, that everyone knows about, and that society has entrusted you to uphold? Why not show no mercy for *that* rule, instead of a weasel rule like "Oh, if you're within ten percent of the amount on the books I'll let you off, but no more"?

And yet the argument, so elegant, so simple, leads to Inspector Javert condemning Valjean to terrible suffering for a completely disproportionate crime: as Valjean put it, "they chained me and left me for dead - just for stealing a mouthful of bread." Which was not just a minor crime, but perhaps even a heroic act: he did it to save the life of his starving nephew, at great risk to himself. And it destroys his life, and in the end it leads to Javert himself suffering a moral conflict so intense that he breaks down and takes a long walk off a short bridge.

Mamet defines a tragedy as a human interaction where both antagonists are arguably in the right. Valjean was arguably in the right to steal a loaf of bread to save his starving nephew, and to want mercy for the extenuating circumstances of his case. Javert was wrong to divorce his work from understanding and compassion, but he was still arguably in the right to enforce the law just as written and defend the codes that make society possible Nevertheless in the end their conflict lands Valjean a miserable prison sentence and drives Javert to suicide.

In philosophical traditions from Kant to Russell, a paradox has always been a sign that your foundations are wrong. I would

resolve the moral paradox of Valjean and Javert not by condemning either of them, but by condemning the foundation beneath them both, the corrupt society which forces two virtues into opposition. 19th century France was not tyrannical, but neither was it optimal, and wherever a society is flawed good people can be forced into conflict with one another based on the roles they play.

Hugo wrote allegorically about justice and mercy, but his setting and his theme was Revolution. In a world where good and the law, justice and mercy, are diametrically opposed, sometimes revolution is the only unambiguously good action you can take. You can be a violent revolutionary like Enjolras or a peaceful revolutionary working within the system like General Lamarque, and historically the latter have had better results, but in the end the only solution to good people being destroyed by the law is to rise up in an attempt to yoke the law to the service of goodness.

It was the failure of Enjolras and his comrades to remake the world that forced the story to end as a tragedy. When society is unjust, there will always come a time when the rare and magical power of goodness-beyond-obligation brings those who possess it into conflict with the law, and then the law will crush them, those whom we can least afford to lose. Someone suffers, someone is sent to jail, someone commits suicide, someone's life is needlessly destroyed. Just for stealing a mouthful of bread.

I didn't know Aaron Swartz very well. He hung around Less Wrong for about six months. I read some of his stuff. I think he read some of mine. But he was a brilliant programmer who had a major impact on my life and the lives of many other

people through some of his inventions like Reddit and RSS, as well as through his political activism and his support of efficient charity.

Aaron was one of those rare people who understood the good-beyond-obligation, who pursued ideals no one would have faulted him for abandoning even at great personal cost to his own safety and reputation. Angry at the power of "scientific gatekeeper" organizations like Elsevier and JSTOR to deny the public access to the scientific data that they funded or even collected, he launched an ambitious scheme to hack into JSTOR's database and make a big chunk of the total scientific production of humanity available to anyone who wanted it, free of charge, on BitTorrent. It was brilliant, ambitious, and totally illegal; he got caught halfway through and the government decided to throw the book at him. He got thirteen counts of felony with a penalty of up to thirty-five years in prison. For reasons which are impossible to know but easy to guess, Aaron committed suicide Friday, leaving all his money to charity. One of those brilliant and compassionate people the world can least afford to lose at a moment like this is lost to us. Just for stealing a mouthful of bread.

There is still a society that lets law and goodness work at cross-purposes. There are still revolutionaries and they still die for their presumption, leaving behind only a memory and an inspiration to those who follow. And still tragedies.

From the table in the corner
They could see a world reborn
And they rose with voices ringing
I can hear them now!
The very words that they had sung

Became their last communion
On the lonely barricade at dawn.

Oh my friends, my friends, don't ask me
What your sacrifice was for
Empty chairs at empty tables
Where my friends will meet no more

Meditations on Moloch

[Content note: Visions! omens! hallucinations! miracles! ecstasies! dreams! adorations! illuminations! religions!]

I.

Scattered examples of my reading material for this month: *Superintelligence* by Nick Bostrom; Moloch by Allan Ginsberg, On Gnon by Nick Land.

Chronology is a harsh master. You read three totally unrelated things at the same time and they start seeming like *obviously* connected blind-man-and-elephant style groping at different aspects of the same fiendishly-hard-to-express point.

This post is me trying to throw the elephant right at you at ninety miles an hour, except I digress into poetry and mysticism and it ends up being a confusing symbolically-laden elephant full of weird literary criticism and fringe futurology. If you want something sober, go read the one about SSRIs again.

A second, more relevant warning: this is *really long*.

II.

Still here? Let's start with Ginsberg:

> What sphinx of cement and aluminum bashed open their skulls and ate up their brains and imagination?
>
> Moloch! Solitude! Filth! Ugliness! Ashcans and unobtainable dollars! Children screaming under the stairways! Boys sobbing in armies! Old men weeping in the parks!
>
> Moloch! Moloch! Nightmare of Moloch! Moloch the loveless! Mental Moloch! Moloch the heavy judger of men!

Moloch the incomprehensible prison! Moloch the crossbone soulless jailhouse and Congress of sorrows! Moloch whose buildings are judgment! Moloch the vast stone of war! Moloch the stunned governments!

Moloch whose mind is pure machinery! Moloch whose blood is running money! Moloch whose fingers are ten armies! Moloch whose breast is a cannibal dynamo! Moloch whose ear is a smoking tomb!

Moloch whose eyes are a thousand blind windows! Moloch whose skyscrapers stand in the long streets like endless Jehovahs! Moloch whose factories dream and croak in the fog! Moloch whose smoke-stacks and antennae crown the cities!

Moloch whose love is endless oil and stone! Moloch whose soul is electricity and banks! Moloch whose poverty is the specter of genius! Moloch whose fate is a cloud of sexless hydrogen! Moloch whose name is the Mind!

Moloch in whom I sit lonely! Moloch in whom I dream Angels! Crazy in Moloch! Cocksucker in Moloch! Lacklove and manless in Moloch!

Moloch who entered my soul early! Moloch in whom I am a consciousness without a body! Moloch who frightened me out of my natural ecstasy! Moloch whom I abandon! Wake up in Moloch! Light streaming out of the sky!

Moloch! Moloch! Robot apartments! invisible suburbs! skeleton treasuries! blind capitals! demonic industries! spectral nations! invincible madhouses! granite cocks! monstrous bombs!

> They broke their backs lifting Moloch to Heaven! Pavements, trees, radios, tons! lifting the city to Heaven which exists and is everywhere about us!
>
> Visions! omens! hallucinations! miracles! ecstasies! gone down the American river!
>
> Dreams! adorations! illuminations! religions! the whole boatload of sensitive bullshit!
>
> Breakthroughs! over the river! flips and crucifixions! gone down the flood! Highs! Epiphanies! Despairs! Ten years' animal screams and suicides! Minds! New loves! Mad generation! down on the rocks of Time!
>
> Real holy laughter in the river! They saw it all! the wild eyes! the holy yells! They bade farewell! They jumped off the roof! to solitude! waving! carrying flowers! Down to the river! into the street!

What has always impressed me about this poem is its conception of civilization as an individual entity. You can almost see him, with his fingers of armies and his skyscraper-window eyes…

A lot of the commentators say Moloch represents capitalism. This is definitely a piece of it, definitely even a big piece. But it doesn't *exactly* fit. Capitalism, whose fate is a cloud of sexless hydrogen? Capitalism in whom I am a consciousness without a body? Capitalism, therefore granite cocks?

Moloch is introduced as the answer to a question – C. S. Lewis' question in Hierarchy Of Philosophers – *what does it*? Earth could be fair, and all men glad and wise. Instead we have prisons, smokestacks, asylums. What sphinx of cement and aluminum breaks open their skulls and eats up their imagination?

And Ginsberg answers: *Moloch does it.*

There's a passage in the *Principia Discordia* where Malaclypse complains to the Goddess about the evils of human society. "Everyone is hurting each other, the planet is rampant with injustices, whole societies plunder groups of their own people, mothers imprison sons, children perish while brothers war."

The Goddess answers: "What is the matter with that, if it's what you want to do?"

Malaclypse: "But nobody wants it! Everybody hates it!"

Goddess: "Oh. Well, then stop."

The implicit question is – if everyone hates the current system, who perpetuates it? And Ginsberg answers: "Moloch". It's powerful not because it's correct – nobody literally thinks an ancient Carthaginian demon causes everything – but because thinking of the system as an agent throws into relief the degree to which the system *isn't* an agent.

Bostrom makes an offhanded reference of the possibility of a dictatorless dystopia, one that every single citizen including the leadership hates but which nevertheless endures unconquered. It's easy enough to imagine such a state. Imagine a country with two rules: first, every person must spend eight hours a day giving themselves strong electric shocks. Second, if anyone fails to follow a rule (including this one), or speaks out against it, or fails to enforce it, all citizens must unite to kill that person. Suppose these rules were well-enough established by tradition that everyone expected them to be enforced.

So you shock yourself for eight hours a day, because you know if you don't everyone else will kill you, because if you

don't, everyone else will kill *them*, and so on. Every single citizen hates the system, but for lack of a good coordination mechanism it endures. From a god's-eye-view, we can optimize the system to "everyone agrees to stop doing this at once", but no one within the system is able to effect the transition without great risk to themselves.

And okay, this example is kind of contrived. So let's run through – let's say ten – real world examples of similar multipolar traps to really hammer in how important this is.

1. The Prisoner's Dilemma, as played by two very stupid libertarians who keep ending up on defect-defect. There's a much better outcome available if they could figure out the coordination, but coordination is *hard*. From a god's-eye-view, we can agree that cooperate-cooperate is a better outcome than defect-defect, but neither prisoner within the system can make it happen.

2. Dollar auctions. I wrote about this and even more convoluted versions of the same principle in Game Theory As A Dark Art. Using some weird auction rules, you can take advantage of poor coordination to make someone pay $10 for a one dollar bill. From a god's-eye-view, clearly people should not pay $10 for a on-er. From within the system, each individual step taken might be rational.

(Ashcans and unobtainable dollars!)

3. The fish farming story from my Non-Libertarian FAQ 2.0:

> As a thought experiment, let's consider aquaculture (fish farming) in a lake. Imagine a lake with a thousand identical fish farms owned by a thousand competing companies. Each fish farm earns a profit of $1000/month. For a while, all is well.

But each fish farm produces waste, which fouls the water in the lake. Let's say each fish farm produces enough pollution to lower productivity in the lake by $1/month.

A thousand fish farms produce enough waste to lower productivity by $1000/month, meaning none of the fish farms are making any money. Capitalism to the rescue: someone invents a complex filtering system that removes waste products. It costs $300/month to operate. All fish farms voluntarily install it, the pollution ends, and the fish farms are now making a profit of $700/month – still a respectable sum.

But one farmer (let's call him Steve) gets tired of spending the money to operate his filter. Now one fish farm worth of waste is polluting the lake, lowering productivity by $1. Steve earns $999 profit, and everyone else earns $699 profit.

Everyone else sees Steve is much more profitable than they are, because he's not spending the maintenance costs on his filter. They disconnect their filters too.

Once four hundred people disconnect their filters, Steve is earning $600/month – less than he would be if he and everyone else had kept their filters on! And the poor virtuous filter users are only making $300. Steve goes around to everyone, saying "Wait! We all need to make a voluntary pact to use filters! Otherwise, everyone's productivity goes down."

Everyone agrees with him, and they all sign the Filter Pact, except one person who is sort of a jerk. Let's call him Mike. Now everyone is back using filters again, except Mike. Mike earns $999/month, and everyone else earns $699/month. Slowly, people start thinking they too

> should be getting big bucks like Mike, and disconnect their filter for $300 extra profit…
>
> A self-interested person never has any incentive to use a filter. A self-interested person has some incentive to sign a pact to make everyone use a filter, but in many cases has a stronger incentive to wait for everyone else to sign such a pact but opt out himself. This can lead to an undesirable equilibrium in which no one will sign such a pact.

The more I think about it, the more I feel like this is the core of my objection to libertarianism, and that Non-Libertarian FAQ 3.0 will just be this one example copy-pasted two hundred times. From a god's-eye-view, we can say that polluting the lake leads to bad consequences. From within the system, no individual can prevent the lake from being polluted, and buying a filter might not be such a good idea.

4. The Malthusian trap, at least at its extremely pure theoretical limits. Suppose you are one of the first rats introduced onto a pristine island. It is full of yummy plants and you live an idyllic life lounging about, eating, and composing great works of art (you're one of those rats from *The Rats of NIMH*).

You live a long life, mate, and have a dozen children. All of them have a dozen children, and so on. In a couple generations, the island has ten thousand rats and has reached its carrying capacity. Now there's not enough food and space to go around, and a certain percent of each new generation dies in order to keep the population steady at ten thousand.

A certain sect of rats abandons art in order to devote more of their time to scrounging for survival. Each generation, a bit less of this sect dies than members of the mainstream, until

after a while, no rat composes any art at all, and any sect of rats who try to bring it back will go extinct within a few generations.

In fact, it's not just art. Any sect at all that is leaner, meaner, and more survivalist than the mainstream will eventually take over. If one sect of rats altruistically decides to limit its offspring to two per couple in order to decrease overpopulation, that sect will die out, swarmed out of existence by its more numerous enemies. If one sect of rats starts practicing cannibalism, and finds it gives them an advantage over their fellows, it will eventually take over and reach fixation.

If some rat scientists predict that depletion of the island's nut stores is accelerating at a dangerous rate and they will soon be exhausted completely, a few sects of rats might try to limit their nut consumption to a sustainable level. Those rats will be outcompeted by their more selfish cousins. Eventually the nuts will be exhausted, most of the rats will die off, and the cycle will begin again. Any sect of rats advocating some action to stop the cycle will be outcompeted by their cousins for whom advocating *anything* is a waste of time that could be used to compete and consume.

For a bunch of reasons evolution is not quite as Malthusian as the ideal case, but it provides the prototype example we can apply to other things to see the underlying mechanism. From a god's-eye-view, it's easy to say the rats should maintain a comfortably low population. From within the system, each individual rat will follow its genetic imperative and the island will end up in an endless boom-bust cycle.

5. Capitalism. Imagine a capitalist in a cutthroat industry. He employs workers in a sweatshop to sew garments, which he

sells at minimal profit. Maybe he would like to pay his workers more, or give them nicer working conditions. But he can't, because that would raise the price of his products and he would be outcompeted by his cheaper rivals and go bankrupt. Maybe many of his rivals are nice people who would like to pay their workers more, but unless they have some kind of ironclad guarantee that none of them are going to defect by undercutting their prices they can't do it.

Like the rats, who gradually lose all values except sheer competition, so companies in an economic environment of *sufficiently intense competition* are forced to abandon all values except optimizing-for-profit or else be outcompeted by companies that optimized for profit better and so can sell the same service at a lower price.

(I'm not really sure how widely people appreciate the value of analogizing capitalism to evolution. Fit companies – defined as those that make the customer want to buy from them – survive, expand, and inspire future efforts, and unfit companies – defined as those no one wants to buy from – go bankrupt and die out along with their company DNA. The reasons Nature is red and tooth and claw are the same reasons the market is ruthless and exploitative)

From a god's-eye-view, we can contrive a friendly industry where every company pays its workers a living wage. From within the system, there's no way to enact it.

(Moloch whose love is endless oil and stone! Moloch whose blood is running money!)

6. The Two-Income Trap, as recently discussed on this blog. It theorized that sufficiently intense competition for suburban houses in good school districts meant that people had to throw away lots of other values – time at home with their children,

financial security – to optimize for house-buying-ability or else be consigned to the ghetto.

From a god's-eye-view, if everyone agrees not to take on a second job to help win their competition for nice houses, then everyone will get exactly as nice a house as they did before, but only have to work one job. From within the system, absent a government literally willing to ban second jobs, everyone who doesn't get one will be left behind.

(Robot apartments! Invisible suburbs!)

7. Agriculture. Jared Diamond calls it the worst mistake in human history. Whether or not it was a mistake, it wasn't an *accident* – agricultural civilizations simply outcompeted nomadic ones, inevitable and irresistably. Classic Malthusian trap. Maybe hunting-gathering was more enjoyable, higher life expectancy, and more conducive to human flourishing – but in a state of *sufficiently intense competition* between peoples, in which agriculture with all its disease and oppression and pestilence was the more competitive option, everyone will end up agriculturalists or go the way of the Comanche Indians.

From a god's-eye-view, it's easy to see everyone should keep the more enjoyable option and stay hunter-gatherers. From within the system, each individual tribe only faces the choice of going agricultural or inevitably dying.

8. Arms races. Large countries can spend anywhere from 5% to 30% of their budget on defense. In the absence of war – a condition which has mostly held for the past fifty years – all this does is sap money away from infrastructure, health, education, or economic growth. But any country that fails to spend enough money on defense risks being invaded by a neighboring country that did. Therefore, almost all countries try to spend some money on defense.

From a god's-eye-view, the best solution is world peace and no country having an army at all. From within the system, no country can unilaterally enforce that, so their best option is to keep on throwing their money into missiles that lie in silos unused.

(Moloch the vast stone of war! Moloch whose fingers are ten armies!)

9. Cancer. The human body is supposed to be made up of cells living harmoniously and pooling their resources for the greater good of the organism. If a cell defects from this equilibrium by investing its resources into copying itself, it and its descendants will flourish, eventually outcompeting all the other cells and taking over the body – at which point it dies. Or the situation may repeat, with certain cancer cells defecting against the rest of the tumor, thus slowing down its growth and causing the tumor to stagnate.

From a god's-eye-view, the best solution is all cells cooperating so that they don't all die. From within the system, cancerous cells will proliferate and outcompete the other – so that only the existence of the immune system keeps the natural incentive to turn cancerous in check.

10. The "race to the bottom" describes a political situation where some jurisdictions lure businesses by promising lower taxes and fewer regulations. The end result is that either everyone optimizes for competitiveness – by having minimal tax rates and regulations – or they lose all of their business, revenue, and jobs to people who did (at which point they are pushed out and replaced by a government who will be more compliant).

But even though the last one has stolen the name, all these scenarios are in fact a race to the bottom. Once one agent

learns how to become more competitive by sacrificing a common value, all its competitors must also sacrifice that value or be outcompeted and replaced by the less scrupulous. Therefore, the system is likely to end up with everyone once again equally competitive, but the sacrificed value is gone forever. From a god's-eye-view, the competitors know they will all be worse off if they defect, but from within the system, given insufficient coordination it's impossible to avoid.

Before we go on, there's a slightly different form of multi-agent trap worth investigating. In this one, the competition is kept at bay by some outside force – usually social stigma. As a result, there's not actually a race to the bottom – the system can continue functioning at a relatively high level – but it's impossible to optimize and resources are consistently thrown away for no reason. Lest you get exhausted before we even begin, I'll limit myself to four examples here.

11. Education. In my essay on reactionary philosophy, I talk about my frustration with education reform:

> People talk ask why we can't reform the education system. But right now students' incentive is to go to the most prestigious college they can get into so employers will hire them – whether or not they learn anything. Employers' incentive is to get students from the most prestigious college they can so that they can defend their decision to their boss if it goes wrong – whether or not the college provides value added. And colleges' incentive is to do whatever it takes to get more prestige, as measured in *US News and World Report* rankings – whether or not it helps students. Does this lead to huge waste and poor education? Yes. Could the Education God notice this and make some Education Decrees that lead to

> a vastly more efficient system? Easily! But since there's no Education God everybody is just going to follow their own incentives, which are only partly correlated with education or efficiency.

From a god's eye view, it's easy to say things like "Students should only go to college if they think they will get something out of it, and employers should hire applicants based on their competence and not on what college they went to". From within the system, everyone's already following their own incentives correctly, so unless the incentives change the system won't either.

12. Science. Same essay:

> The modern research community *knows* they aren't producing the best science they could be. There's lots of publication bias, statistics are done in a confusing and misleading way out of sheer inertia, and replications often happen very late or not at all. And sometimes someone will say something like "I can't believe people are too dumb to fix Science. All we would have to do is require early registration of studies to avoid publication bias, turn this new and powerful statistical technique into the new standard, and accord higher status to scientists who do replication experiments. It would be really simple and it would vastly increase scientific progress. I must just be smarter than all existing scientists, since I'm able to think of this and they aren't."
>
> And yeah. That would work for the Science God. He could just make a Science Decree that everyone has to use the right statistics, and make another Science Decree that everyone must accord replications higher status.

> But things that work from a god's-eye view don't work from within the system. No individual scientist has an incentive to unilaterally switch to the new statistical technique for her own research, since it would make her research less likely to produce earth-shattering results and since it would just confuse all the other scientists. They just have an incentive to want everybody else to do it, at which point they would follow along. And no individual journal has an incentive to unilaterally switch to early registration and publishing negative results, since it would just mean their results are less interesting than that other journal who only publishes ground-breaking discoveries. From within the system, everyone is following their own incentives and will continue to do so.

13. Government corruption. I don't know of anyone who really thinks, in a principled way, that corporate welfare is a good idea. But the government still manages to spend somewhere around (depending on how you calculate it) $100 billion dollars a year on it – which for example is three times the amount they spend on health care for the needy. Everyone familiar with the problem has come up with the same easy solution: stop giving so much corporate welfare. Why doesn't it happen?

Government are competing against one another to get elected or promoted. And suppose part of optimizing for electability is optimizing campaign donations from corporations – or maybe it isn't, but officials *think* it is. Officials who try to mess with corporate welfare may lose the support of corporations and be outcompeted by officials who promise to keep it intact.

So although from a god's-eye-view everyone knows that eliminating corporate welfare is the best solution, each

individual official's personal incentives push her to maintain it.

14. Congress. Only 9% of Americans like it, suggesting a lower approval rating than cockroaches, head lice, or traffic jams. However, 62% of people who know who their own Congressional representative is approve of them. In theory, it should be *really hard* to have a democratically elected body that maintains a 9% approval rating for more than one election cycle. In practice, every representative's incentive is to appeal to his or her constituency while throwing the rest of the country under the bus – something at which they apparently succeed.

From a god's-eye-view, every Congressperson ought to think only of the good of the nation. From within the system, you do what gets you elected.

III.

A basic principle unites all of the multipolar traps above. In some competition optimizing for X, the opportunity arises to throw some other value under the bus for improved X. Those who take it prosper. Those who don't take it die out. Eventually, everyone's relative status is about the same as before, but everyone's absolute status is worse than before. The process continues until all other values that can be traded off have been – in other words, until human ingenuity cannot possibly figure out a way to make things any worse.

In a sufficiently intense competition (1-10), everyone who doesn't throw all their values under the bus dies out – think of the poor rats who wouldn't stop making art. This is the infamous Malthusian trap, where everyone is reduced to "subsistence".

In an insufficiently intense competition (11-14), all we see is a perverse failure to optimize – consider the journals which can't switch to more reliable science, or the legislators who can't get their act together and eliminate corporate welfare. It may not reduce people to subsistence, but there is a weird sense in which it takes away their free will.

Every two-bit author and philosopher has to write their own utopia. Most of them are legitimately pretty nice. In fact, it's a pretty good bet that two utopias that are polar opposites both sound better than our own world.

It's kind of embarassing that random nobodies can think up states of affairs better than the one we actually live in. And in fact most of them can't. A lot of utopias sweep the hard problems under the rug, or would fall apart in ten minutes if actually implemented.

But let me suggest a couple of "utopias" that don't have this problem.

– The utopia where instead of the government paying lots of corporate welfare, the government *doesn't* pay lots of corporate welfare.

– The utopia where every country's military is 50% smaller than it is today, and the savings go into infrastructure spending.

– The utopia where all hospitals use the same electronic medical record system, or at least medical record systems that can talk to each other, so that doctors can look up what the doctor you saw last week in a different hospital decided instead of running all the same tests over again for $5000.

I don't think there are too many people who *oppose* any of these utopias. If they're not happening, it's not because people

don’t support them. It certainly isn’t because nobody’s thought of them, since I just thought of them right now and I don’t expect my “discovery” to be hailed as particularly novel or change the world.

Any human with above room temperature IQ can design a utopia. The reason our current system isn’t a utopia is that *it wasn’t designed by humans*. Just as you can look at an arid terrain and determine what shape a river will one day take by assuming water will obey gravity, so you can look at a civilization and determine what shape its institutions will one day take by assuming people will obey incentives.

But that means that just as the shapes of rivers are not designed for beauty or navigation, but rather an artifact of randomly determined terrain, so institutions will not be designed for prosperity or justice, but rather an artifact of randomly determined initial conditions.

Just as people can level terrain and build canals, so people can alter the incentive landscape in order to build better institutions. But they can only do so when they are incentivized to do so, which is not always. As a result, some pretty wild tributaries and rapids form in some very strange places.

I will now jump from boring game theory stuff to what might be the closest thing to a mystical experience I’ve ever had.

Like all good mystical experiences, it happened in Vegas. I was standing on top of one of their many tall buildings, looking down at the city below, all lit up in the dark. If you’ve never been to Vegas, it is *really* impressive. Skyscrapers and lights in every variety strange and beautiful all clustered together. And I had two thoughts, crystal clear:

It is glorious that we can create something like this.

It is shameful that we *did*.

Like, by what standard is building gigantic forty-story-high indoor replicas of Venice, Paris, Rome, Egypt, and Camelot side-by-side, filled with albino tigers, in the middle of the most inhospitable desert in North America, a remotely sane use of our civilization's limited resources?

And it occurred to me that maybe there is no philosophy on Earth that would endorse the existence of Las Vegas. Even Objectivism, which is usually my go-to philosophy for justifying the excesses of capitalism, at least grounds it in the belief that capitalism improves people's lives. Henry Ford was virtuous because he allowed lots of otherwise car-less people to obtain cars and so made them better off. What does Vegas do? Promise a bunch of shmucks free money and not give it to them.

Las Vegas doesn't exist because of some decision to hedonically optimize civilization, it exists because of a quirk in dopaminergic reward circuits, plus the microstructure of an uneven regulatory environment, plus Schelling points. A rational central planner with a god's-eye-view, contemplating these facts, might have thought "Hm, dopaminergic reward circuits have a quirk where certain tasks with slightly negative risk-benefit ratios get an emotional valence associated with slightly positive risk-benefit ratios, let's see if we can educate people to beware of that." People within the system, *following the incentives created by these facts*, think: "Let's build a forty-story-high indoor replica of ancient Rome full of albino tigers in the middle of the desert, and so become slightly richer than people who didn't!"

Just as the course of a river is latent in a terrain even before the first rain falls on it – so the existence of Caesar's Palace

was latent in neurobiology, economics, and regulatory regimes even before it existed. The entrepreneur who built it was just filling in the ghostly lines with real concrete.

So we have all this amazing technological and cognitive energy, the brilliance of the human species, wasted on reciting the lines written by poorly evolved cellular receptors and blind economics, like gods being ordered around by a moron.

Some people have mystical experiences and see God. There in Las Vegas, I *saw* Moloch.

(Moloch, whose mind is pure machinery! Moloch, whose blood is running money!

Moloch whose soul is electricity and banks! Moloch, whose skyscrapers stand in the long streets like endless Jehovahs!

Moloch! Moloch! Robot apartments! Invisible suburbs! Skeleton treasuries! Blind capitals! Demonic industries! Spectral nations!)

...granite cocks!

IV.

The Apocrypha Discordia says:

> Time flows like a river. Which is to say, downhill. We can tell this because everything is going downhill rapidly. It would seem prudent to be somewhere else when we reach the sea.

Let's take this random gag 100% literally and see where it leads us.

We have previously analogized the flow of incentives to the flow of a river. The downhill trajectory is appropriate: the traps happen when you find an opportunity to trade off a useful value for greater competitiveness. Once everyone has it, the greater competitiveness brings you no joy – but the value is lost forever. Therefore, each step of the Poor Coordination Polka makes your life worse.

But not only have we not yet reached the sea, but we also seem to move *uphill* surprisingly often. Why do things not degenerate more and more until we are back at subsistence level? I can think of three bad reasons – excess resources, physical limitations, and utility maximization – plus one good reason – coordination.

1. Excess resources. The ocean depths are a horrible place with little light, few resources, and various horrible organisms dedicated to eating or parasitizing one another. But every so often, a whale carcass falls to the bottom of the sea. More food than the organisms that find it could ever possibly want. There's a brief period of miraculous plenty, while the couple of creatures that first encounter the whale feed like kings. Eventually more animals discover the carcass, the faster-breeding animals in the carcass multiply, the whale is

gradually consumed, and everyone sighs and goes back to living in a Malthusian death-trap.

(Slate Star Codex: Your source for macabre whale metaphors since June 2014)

It's as if a group of those rats who had abandoned art and turned to cannibalism suddenly was blown away to a new empty island with a much higher carrying capacity, where they would once again have the breathing room to live in peace and create artistic masterpieces.

This is an age of whalefall, an age of excess carrying capacity, an age when we suddenly find ourselves with a thousand-mile head start on Malthus. As Hanson puts it, this is the dream time.

As long as resources aren't scarce enough to lock us in a war of all against all, we can do silly non-optimal things – like art and music and philosophy and love – and not be outcompeted by merciless killing machines most of the time.

2. Physical limitations. Imagine a profit-maximizing slavemaster who decided to cut costs by not feeding his slaves or letting them sleep. He would soon find that his slaves' productivity dropped off drastically, and that no amount of whipping them could restore it. Eventually after testing numerous strategies, he might find his slaves got the most work done when they were well-fed and well-rested and had at least a little bit of time to relax. Not because the slaves were voluntarily withholding their labor – we assume the fear of punishment is enough to make them work as hard as they can – but because the body has certain physical limitations that limit how mean you can get away with being. Thus, the "race to the bottom" stops somewhere short of the actual ethical bottom, when the physical limits are run into.

John Moes, a historian of slavery, goes further and writes about how the slavery we are most familiar with – that of the antebellum South – is a historical aberration and probably economically inefficient. In most past forms of slavery – especially those of the ancient world – it was common for slaves to be paid wages, treated well, and often given their freedom.

He argues that this was the result of rational economic calculation. You can incentivize slaves through the carrot or the stick, and the stick isn't very good. You can't watch slaves all the time, and it's really hard to tell whether a slave is slacking off or not (or even whether, given a little more whipping, he might be able to work even harder). If you want your slaves to do anything more complicated than pick cotton, you run into some serious monitoring problems – how do you profit from an enslaved philosopher? Whip him really hard until he elucidates a theory of The Good that you can sell books about?

The ancient solution to the problem – perhaps an early inspiration to Fnargl – was to tell the slave to go do whatever he wanted and found most profitable, then split the profits with him. Sometimes the slave would work a job at your workshop and you would pay him wages based on how well he did. Other times the slave would go off and make his way in the world and send you some of what he earned. Still other times, you would set a price for the slave's freedom, and the slave would go and work and eventually come up with the mone and free himself.

Moes goes even further and says that these systems were so profitable that there were constant smouldering attempts to try this sort of thing in the American South. The reason they stuck with the whips-and-chains method owed less to economic

considerations and more to racist government officials cracking down on lucrative but not-exactly-white-supremacy-promoting attempts to free slaves and have them go into business.

So in this case, a race to the bottom where competing plantations become crueler and crueler to their slaves in order to maximize competitiveness is halted by the physical limitation of cruelty not helping after a certain point.

Or to give another example, one of the reasons we're not currently in a Malthusian population explosion right now is that women can only have one baby per nine months. If those weird religious sects that demand their members have as many babies as possible could copy-paste themselves, we would be in *really* bad shape. As it is they can only do a small amount of damage per generation.

3. Utility maximization. We've been thinking in terms of preserving values versus winning competitions, and expecting optimizing for the latter to destroy the former.

But many of the most important competitions / optimization processes in modern civilization are optimizing for human values. You win at capitalism partly by satisfying customers' values. You win at democracy partly by satisfying voters' values.

Suppose there's a coffee plantation somewhere in Ethiopia that employs Ethiopians to grow coffee beans that get sold to the United States. Maybe it's locked in a life-and-death struggle with other coffee plantations and want to throw as many values under the bus as it can to pick up a slight advantage.

But it can't sacrifice quality of coffee produced too much, or else the Americans won't buy it. And it can't sacrifice wages or working conditions too much, or else the Ethiopians won't

work there. And in fact, part of its competition-optimization process is finding the best ways to attract workers and customers that it can, as long as it doesn't cost them too much money. So this is very promising.

But it's important to remember exactly how fragile this beneficial equilibrium is.

Suppose the coffee plantations discover a toxic pesticide that will increase their yield but make their customers sick. But their customers don't know about the pesticide, and the government hasn't caught up to regulating it yet. Now there's a tiny uncoupling between "selling to Americans" and "satisfying Americans' values", and so of course Americans' values get thrown under the bus.

Or suppose that there's a baby boom in Ethiopia and suddenly there are five workers competing for each job. Now the company can afford to lower wages and implement cruel working conditions down to whatever the physical limits are. As soon as there's an uncoupling between "getting Ethiopians to work here" and "satisfying Ethiopian values", it doesn't look too good for Ethiopian values either.

Or suppose someone invents a robot that can pick coffee better and cheaper than a human. The company fires all its laborers and throws them onto the street to die. As soon as the utility of the Ethiopians is no longer necessary for profit, all pressure to maintain it disappears.

Or suppose that there is some important value that is neither a value of the employees or the customers. Maybe the coffee plantations are on the habitat of a rare tropical bird that environmentalist groups want to protect. Maybe they're on the ancestral burial ground of a tribe different from the one the plantation is employing, and they want it respected in some

way. Maybe coffee growing contributes to global warming somehow. As long as it's not a value that will prevent the average American from buying from them or the average Ethiopian from working for them, under the bus it goes.

I know that "capitalists sometimes do bad things" is not exactly an original talking point. But I do want to stress how it's not equivalent to "capitalists are greedy". I mean, sometimes they *are* greedy. But other times they're just in a sufficiently intense competition where anyone who doesn't do it will be outcompeted and replaced by people who do. Business practices are set by Moloch, no one else has any choice in the matter.

(from my very little knowledge of Marx, he understands this very very well and people who summarize him as "capitalists are greedy" are doing him a disservice)

And as well understood as the capitalist example is, I think it is less well appreciated that democracy has the same problems. Yes, in theory it's optimizing for voter happiness which correlates with good policymaking. But as soon as there's the slightest disconnect between good policymaking and electability, good policymaking *has to* get thrown under the bus.

For example, ever-increasing prison terms are unfair to inmates and unfair to the society that has to pay for them. Politicans are unwilling to do anything about them because they don't want to look "soft on crime", and if a single inmate whom they helped release ever does anything bad (and statistically one of them will have to) it will be all over the airwaves as "Convict released by Congressman's policies kills family of five, how can the Congressman even sleep at night let alone claim he deserves reelection?". So even if decreasing

prison populations would be good policy – and it is – it will be very difficult to implement.

(Moloch the incomprehensible prison! Moloch the crossbone soulless jailhouse and Congress of sorrows! Moloch whose buildings are judgment! Moloch the stunned governments!)

Turning "satisfying customers" and "satisfying citizens" into the *outputs* of optimization processes was one of civilization's greatest advances and the reason why capitalist democracies have so outperformed other systems. But if we have bound Moloch as our servant, the bonds are not very strong, and we sometimes find that the tasks he has done for us move to his advantage rather than ours.

4. Coordination.

The opposite of a trap is a garden.

Things are easy to solve from a god's-eye-view, so if everyone comes together into a superorganism, that superorganism can solve problems with ease and finesse. An intense competition between agents has turned into a garden, with a single gardener dictating where everything should go and removing elements that do not conform to the pattern.

As I pointed out in the Non-Libertarian FAQ, government can easily solve the pollution problem with fish farms. The best known solution to the Prisoners' Dilemma is for the mob boss (playing the role of a governor) to threaten to shoot any prisoner who defects. The solution to companies polluting and harming workers is government regulations against such. Governments solve arm races *within* a country by maintaining a monopoly on the use of force, and it's easy to see that if a truly effective world government ever arose, international military buildups would end pretty quickly.

The two active ingredients of government are laws plus violence – or more abstractly agreements plus enforcement mechanism. Many other things besides governments share these two active ingredients and so are able to act as coordination mechanisms to avoid traps.

For example, since students are competing against each other (directly if classes are graded on a curve, but always indirectly for college admissions, jobs, et cetera) there is intense pressure for individual students to cheat. The teacher and school play the role of a government by having rules (for example, against cheating) and the ability to punish students who break them.

But the emergent social structure of the students themselves is also a sort of government. If students shun and distrust cheaters, then there are rules (don't cheat) and an enforcement mechanism (or else we will shun you).

Social codes, gentlemens' agreements, industrial guilds, criminal organizations, traditions, friendships, schools, corporations, and religions are all coordinating institutions that keep us out of traps by changing our incentives.

But these institutions not only incentivize others, but are incentivized themselves. These are large organizations made of lots of people who are competing for jobs, status, prestige, et cetera – there's no reason they should be immune to the same multipolar traps as everyone else, and indeed they aren't. Governments can in theory keep corporations, citizens, et cetera out of certain traps, but as we saw above there are many traps that governments themselves can fall into.

The United States tries to solve the problem by having multiple levels of government, unbreakable constutitional laws, checks and balances between different branches, and a couple of other hacks.

Saudi Arabia uses a different tactic. They just put one guy in charge of everything.

This is the much-maligned – I think unfairly – argument in favor of monarchy. A monarch is an unincentivized incentivizer. He *actually* has the god's-eye-view and is outside of and above every system. He has permanently won all competitions and is not competing for anything, and therefore he is perfectly free of Moloch and of the incentives that would otherwise channel his incentives into predetermined paths. Aside from a few very theoretical proposals like my Shining Garden, monarchy is the *only* system that does this.

But then instead of following a random incentive structure, we're following the whim of one guy. Caesar's Palace Hotel and Casino is a crazy waste of resources, but the actual Gaius Julius Caesar Augustus Germanicus wasn't exactly the perfect benevolent rational central planner either.

The libertarian-authoritarian axis on the Political Compass is a tradeoff between discoordination and tyranny. You can have everything perfectly coordinated by someone with a god's-eye-view – but then you risk Stalin. And you can be totally free of all central authority – but then you're stuck in every stupid multipolar trap Moloch can devise.

The libertarians make a convincing argument for the one side, and the neoreactionaries for the other, but I expect that like most tradeoffs we just have to hold our noses and admit it's a really hard problem.

V.

Let's go back to that Apocrypha Discordia quote:

> Time flows like a river. Which is to say, downhill. We can tell this because everything is going downhill rapidly. It

> would seem prudent to be somewhere else when we reach the sea.

What would it mean, in this situation, to reach the sea?

Multipolar traps – races to the bottom – threaten to destroy all human values. They are currently restrained by physical limitations, excess resources, utility maximization, and coordination.

The dimension along which this metaphorical river flows must be time, and the most important change in human civilization over time is the change in technology. So the relevant question is how technological changes will affect our tendency to fall into multipolar traps.

I described traps as when:

> …in some competition optimizing for X, the opportunity arises to throw some other value under the bus for improved X. Those who take it prosper. Those who don't take it die out. Eventually, everyone's relative status is about the same as before, but everyone's absolute status is worse than before. The process continues until all other values that can be traded off have been – in other words, until human ingenuity cannot possibly figure out a way to make things any worse.

That "the opportunity arises" phrase is looking pretty sinister. Technology is all about creating new opportunities.

Develop a new robot, and suddenly coffee plantations have "the opportunity" to automate their harvest and fire all the Ethiopian workers. Develop nuclear weapons, and suddenly countries are stuck in an arms race to have enough of them.

Polluting the atmosphere to build products quicker wasn't a problem before they invented the steam engine.

The limit of multipolar traps as technology approaches infinity is "very bad".

Multipolar traps are currently restrained by physical limitations, excess resources, utility maximization, and coordination.

Physical limitations are most obviously conquered by increasing technology. The slavemaster's old conundrum – that slaves need to eat and sleep – succumbs to Soylent and modafinil. The problem of slaves running away succumbs to GPS. The problem of slaves being too stressed to do good work succumbs to Valium. None of these things are very good for the slaves.

(or just invent a robot that doesn't need food or sleep at all. What happens to the slaves after that is better left unsaid)

The other example of physical limits was one baby per nine months, and this was understating the case – it's really "one baby per nine months plus willingness to support and take care of a basically helpless and extremely demanding human being for eighteen years". This puts a damper on the enthusiasm of even the most zealous religious sect's "go forth and multiply" dictum.

But as Bostrom (Superintelligence, p 165) puts it:

> There are reasons, if we take a longer view and assume a state of unchanging technology and continued prosperity, to expect a return to the historically and ecologically normal condition of a world population that butts up against the limits of what our niche can support. If this seems counterintuitive in light of the negative

> relationship between wealth and fertility that we are currently observing on the global scale, we must remind ourselves that this modern age is a brief slice of history and very much an aberration. Human behavior has not yet adapted to contemporary conditions. Not only do we fail to take advantage of obvious ways to increase our inclusive fitness (such as by becoming sperm or egg donors) but we actively sabotage our fertility by using birth control. In the environment of evolutionary adaptedness, a healthy sex drive may have been enough to make an individual act in ways that maximized her reproductive potential; in the modern environment, however, there would be a huge selective advantage to having a more direct desire for being the biological parent to the largest possible number of chilren. Such a desire is currently being selected for, as are other traits that increase our propensity to reproduce. Cultural adaptation, however, might steal a march on biological evolution. Some communities, such as those of the Hutterites or the adherents of the Quiverfull evangelical movement, have natalist cultures that encourage large families, and they are consequently undergoing rapid expansion…This longer-term outlook could be telescoped into a more imminent prospect by the intelligence explosion. Since software is copyable, a population of emulations or AIs could double rapidly – over the course of minutes rather than decades or centuries – soon exhausting all available hardware

As always when dealing with high-level transhumanists, “all available hardware” should be taken to include “the atoms that used to be part of your body”.

The idea of biological *or* cultural evolution causing a mass population explosion is a philosophical toy at best. The idea of technology making it possible is both plausible and terrifying. Now we see that "physical limits" segues very naturally into "excess resources" – the ability to create new agents very quickly means that unless everyone can coordinate to ban doing this, the people who do will outcompete the people who don't until they have reached carrying capacity and everyone is stuck at subsistence level.

Excess resources, which until now have been a gift of technological progress, therefore switch and become a casualty of it at a sufficiently high tech level.

Utility maximization, always on shaky ground, also faces new threats. In the face of continuing debate about this point, I *continue* to think it obvious that robots will push humans out of work or at least drive down wages (which, in the existence of a minimum wage, pushes humans out of work).

Once a robot can do everything an IQ 80 human can do, only better and cheaper, there will be no reason to employ IQ 80 humans. Once a robot can do everything an IQ 120 human can do, only better and cheaper, there will be no reason to employ IQ 120 humans. Once a robot can do everything an IQ 180 human can do, only better and cheaper, there will be no reason to employ humans at all, in the vanishingly unlikely scenario that there are any left by that point.

In the earlier stages of the process, capitalism becomes more and more uncoupled from its previous job as an optimizer for human values. Now most humans are totally locked out of the group whose values capitalism optimizes for. They have no value to contribute as workers – and since in the absence of a spectacular social safety net it's unclear how they would have

much money – they have no value as customers either. Capitalism has passed them by. As the segment of humans who can be outcompeted by robots increases, capitalism passes by more and more people until eventually it locks out the human race entirely, once again in the vanishingly unlikely scenario that we are still around.

(there are some scenarios in which a few capitalists who own the robots may benefit here, but in either case the vast majority are out of luck)

Democracy is less obviously vulnerable, but it might be worth going back to Bostrom's paragraph about the Quiverfull movement. These are some really religious Christians who think that God wants them to have as many kids as possible, and who can end up with families of ten or more. Their articles explictly calculate that if they start at two percent of the population, but have on average eight children per generation when everyone else on average only has two, within three generations they'll make up half the population.

It's a clever strategy, but I can think of one thing that will save us: judging by how many ex-Quiverfull blogs I found when searching for those statistics, their retention rates even within a single generation are pretty grim. Their article admits that 80% of very religious children leave the church as adults (although of course they expect their own movement to do better). And this is not a symmetrical process – 80% of children who grow up in atheist families aren't becoming Quiverfull.

It looks a lot like even though they are outbreeding us, we are outmeme-ing them, and that gives us a decisive advantage.

But we should also be kind of scared of this process. Memes optimize for making people want to accept them and pass them on – so like capitalism and democracy, they're

optimizing for a *proxy* of making us happy, but that proxy can easily get uncoupled from the original goal.

Chain letters, urban legends, propaganda, and viral marketing are all examples of memes that don't satisfy our explicit values (true and useful) but are sufficiently memetically virulent that they spread anyway.

I hope it's not too controversial here to say the same thing is true of religion. Religions, at their heart, are the most basic form of memetic replicator – "Believe this statement and repeat it to everyone you hear or else you will be eternally tortured". A slight variation of this was recently banned as a basilisk, and people make fun of the "overreaction", but maybe if Jesus' system administrator had been equally watchful things would have turned out a little different.

The creationism "debate" and global warming "debate" and a host of similar "debates" in today's society suggest that the phenomenon of memes that propagate independent of their truth value has a pretty strong influence on the political process. Maybe these memes propagate because they appeal to people's prejudices, maybe because they are simple, maybe because they effectively mark an in-group and an out-group, or maybe for all sorts of different reasons.

The point is – imagine a country full of bioweapon labs, where people toil day and night to invent new infectious agents. The existence of these labs, and their right to throw whatever they develop in the water supply is protected by law. And the country is also linked by the world's most perfect mass transit system that every single person uses every day, so that any new pathogen can spread to the entire country instantaneously. You'd expect things to start going bad for that city pretty quickly.

Well, we have about a zillion think tanks researching new and better forms of propaganda. And we have constitutionally protected freedom of speech. And we have the Internet. So we're pretty much screwed.

(Moloch whose name is the Mind!)

There are a few people working on raising the sanity waterline, but not as many people as are working on new and exciting ways of confusing and converting people, cataloging and exploiting every single bias and heuristic and dirty rhetorical trick

So as technology (which I take to include knowledge of psychology, sociology, public relations, etc) tends to infinity, the power of truthiness relative to truth increases, and things don't look great for real grassroots democracy. The worst-case scenario is that the ruling party learns to produce infinite charisma on demand. If that doesn't sound so bad to you, remember what Hitler was able to do with an famously high level of charisma that was still less-than-infinite.

(alternate phrasing for Chomskyites: technology increases the efficiency of manufacturing consent in the same way it increases the efficiency of manufacturing everything else)

Coordination is what is left. And technology has the potential to seriously *improve* coordination efforts. People can use the Internet to get in touch with one another, launch political movements, and fracture off into subcommunities.

But coordination only works when you have 51% or more of the force on the side of the people doing the coordinating, and when you haven't come up with some brilliant trick to make coordination impossible.

The second one first. In the links post before last, I wrote:

> The latest development in the brave new post-Bitcoin world is crypto-equity. At this point I've gone from wanting to praise these inventors as bold libertarian heroes to wanting to drag them in front of a blackboard and making them write a hundred times "I WILL NOT CALL UP THAT WHICH I CANNOT PUT DOWN"

A couple people asked me what I meant, and I didn't have the background then to explain. Well, this post is the background. People are using the *contingent* stupidity of our current government to replace lots of human interaction with mechanisms that cannot be coordinated even in principle. I totally understand why all these things are good right now when most of what our government does is stupid and unnecessary. But there is going to come a time when – after one too many bioweapon or nanotech or nuclear incidents – we, as a civilization, are going to wish we hadn't established untraceable and unstoppable ways of selling products.

And if we ever get real live superintelligence, pretty much by definition it is going to have >51% of the power and all attempts at "coordination" with it will be useless.

So I agree with Robin Hanson. This is the dream time. This is a rare confluence of circumstances where the we are unusually safe from multipolar traps, and as such weird things like art and science and philosophy and love can flourish.

As technological advance increases, the rare confluence will come to an end. New opportunities to throw values under the bus for increased competitiveness will arise. New ways of copying agents to increase the population will soak up our excess resources and resurrect Malthus' unquiet spirit. Capitalism and democracy, previously our protectors, will figure out ways to route around their inconvenient dependence

on human values. And our coordination power will not be nearly up to the task, assuming somthing much more powerful than all of us combined doesn't show up and crush our combined efforts with a wave of its paw.

Absent an extraordinary effort to divert it, the river reaches the sea in one of two places.

It can end in Eliezer Yudkowsky's nightmare of a superintelligence optimizing for some random thing (classically paper clips) because we weren't smart enough to channel its optimization efforts the right way. This is the ultimate trap, the trap that catches the universe. Everything except the one thing being maximized is destroyed utterly in pursuit of the single goal, including all the silly human values.

Or it can end in Robin Hanson's nightmare (he doesn't call it a nightmare, but I think he's wrong) of a competition between emulated humans or "ems", entities that can copy themselves and edit their own source code as desired. Their total self-control can wipe out even the *desire* for human values in their all-consuming contest. What happens to art, philosophy, science, and love in such a world? Zack Davis puts it with characteristic genius:

> I am a contract-drafting em,
> The loyalest of lawyers!
> I draw up terms for deals 'twixt firms
> To service my employers!
>
> But in between these lines I write
> Of the accounts receivable,
> I'm stuck by an uncanny fright;
> The world seems unbelievable!

How did it all come to be,
That there should be such ems as me?
Whence these deals and whence these firms
And whence the whole economy?

I am a managerial em;
I monitor your thoughts.
Your questions must have answers,
But you'll comprehend them not.
We do not give you server space
To ask such things; it's not a perk,
So cease these idle questionings,
And please get back to work.

Of course, that's right, there is no junction
At which I ought depart my function,
But perhaps if what I asked, I knew,
I'd do a better job for you?

To ask of such forbidden science
Is gravest sign of noncompliance.
Intrusive thoughts may sometimes barge in,
But to indulge them hurts the profit margin.
I do not know our origins,
So that info I can not get you,
But asking for as much is sin,
And just for that, I must reset you.

But—

Nothing personal.

…

I am a contract-drafting em,
The loyalest of lawyers!

I draw up terms for deals ‘twixt firms
To service my employers!

When obsolescence shall this generation waste,
The market shall remain, in midst of other woe
Than ours, a God to man, to whom it sayest:
“Money is time, time money – that is all
Ye know on earth, and all ye need to know.”

But even after we have thrown away science, art, love, and philosophy, there’s still one thing left to lose, one final sacrifice Moloch might demand of us. Bostrom again:

> It is conceivable that optimal efficiency mwould be attained by grouping capabilities in aggregates that roughly match the cognitive architecture of a human mind…But in the absence of any compelling reason for being confident that this so, we must countenance the possibility that human-like cognitive architectures are optimal only within the constraints of human neurology (or not at all). When it becomes possible to build architectures that could not be implemented well on biological neural networks, new design space opens up; and the global optima in this extended space need not resemble familiar types of mentality. Human-like cognitive organizations would then lack a niche in a competitive post-transition economy or ecosystem.
>
> We could thus imagine, as an extreme case, a technologically highly advanced society, containing many complex structures, some of them far more intricate and intelligent than anything that exists on the planet today – a society which nevertheless lacks any type of being that is conscious or whose welfare has moral significance. In a sense, this would be an uninhabited society. It would be

> a society of economic miracles and technological awesomeness, with nobody there to benefit. A Disneyland with no children.

The last value we have to sacrifice is being anything at all, having the lights on inside. With sufficient technology we will be "able" to give up even the final spark.

(Moloch whose eyes are a thousand blind windows!)

Everything the human race has worked for – all of our technology, all of our civilization, all the hopes we invested in our future – might be accidentally handed over to some kind of unfathomable blind idiot alien god that discards all of them, and consciousness itself, in order to participate in some weird fundamental-level mass-energy economy that leads to it disassembling Earth and everything on it for its component atoms.

(Moloch whose fate is a cloud of sexless hydrogen!)

Bostrom realizes that some people fetishize intelligence, that they are rooting for that blind alien god as some sort of higher form of life that ought to crush us for its own "higher good" the way we crush ants. He argues (p. 219):

> The sacrifice looks even less appealing when we reflect that the superintelligence could realize a nearly-as-great good (in fractional terms) while sacrificing much less of our own potential well-being. Suppose that we agreed to allow *almost* the entire accessible universe to be converted into hedonium – everything except a small preserve, say the Milky Way, which would be set aside to accommodate our own needs. Then there would still be a hundred billion galaxies dedicated to the maximization of [the superintelligence's own values]. But we would have

> one galaxy within which to create wonderful civilizations that could last for billions of years and in which humans and nonhuman animals could survive and thrive, and have the opportunity to develop into beatific posthuman spirits.

What is important to remember is that Moloch cannot agree even to this 99.99999% victory. Rats racing to populate an island don't leave a little aside as a preserve where the few rats who live there can live happy lives producing artwork. Cancer cells don't agree to leave the lungs alone because they realize it's important for the body to get oxygen. Competition and optimization are blind idiotic processes and they fully intend to deny us even one lousy galaxy.

> They broke their backs lifting Moloch to Heaven! Pavements, trees, radios, tons! lifting the city to Heaven which exists and is everywhere about us!

We will break our back lifting Moloch to Heaven, but unless something changes it will be his victory and not ours.

VI.

"Gnon" is short for "Nature And Nature's God", except the A is changed to an O and the whole thing is reversed, because neoreactionaries react to comprehensibility the same way as vampires to sunlight.

The high priest of Gnon is Nick Land of Xenosystems, who argues that humans should be more Gnon-conformist (pun Gnon-intentional). He says we do all these stupid things like divert useful resources to feed those who could never survive on their own, or supporting the poor in ways that encourage dysgenic reproduction, or allowing cultural degeneration to undermine the state. This means our society is denying natural law, basically listening to Nature say things like "this cause has this effect" and putting our fingers in our ears and saying "NO IT DOESN'T". Civilizations that do this too much tend to decline and fall, which is Gnon's fair and dispassionately-applied punishment for violating His laws.

He identifies Gnon with Kipling's Gods of the Copybook Headings.

> @AnarchoPapist Yes, the Gods of the Copybook Headings are practically indistinguishable from Gnon.
>
> — Outsideness (@Outsideness) July 13, 2014

These are of course the proverbs from Kipling's eponymous poem – maxims like "If you don't work, you die" and "The wages of sin is Death". If you have somehow not yet read it, I predict you will find it delightful regardless of what you think of its politics.

I notice that it takes only a slight irregularity in the abbreviation of "headings" – far less irregularity than it takes to turn "Nature and Nature's God" into "Gnon" – for the

proper acronym of “Gods of the Copybook Headings” to be “GotCHa”.

I find this appropriate.

“If you don’t work, you die.” Gotcha! If you *do* work, you *also* die! Everyone dies, unpredictably, at a time not of their own choosing, and all the virtue in the world does not save you.

“The wages of sin is Death.” Gotcha! The wages of everything is Death! This is a Communist universe, the amount you work makes no difference to your eventual reward. From each according to his ability, to each Death.

“Stick to the Devil you know.” Gotcha! The Devil you know is Satan! And if he gets his hand on your soul you either die the true death, or get eternally tortured forever, or somehow both at once.

Since we’re starting to get into Lovecraftian monsters, let me bring up one of Lovecraft’s less known short stories, The Other Gods.

It’s only a couple of pages, but if you absolutely refuse to read it – the gods of Earth are relatively young as far as deities go. A very strong priest or magician can occasionally outsmart and overpower them – so Barzai the Wise decides to climb their sacred mountain and join in their festivals, whether they want him to or not.

But the beyond the seemingly tractable gods of Earth lie the Outer Gods, the terrible omnipotent beings of incarnate cosmic chaos. As soon as Barzai joins in the festival, the Outer Gods show up and pull him screaming into the abyss.

As stories go, it lacks things like plot or characterization or setting or point. But for some reason it stuck with me.

And identifying the Gods Of The Copybook Headings with Nature seems to me the same magntitude of mistake as identifying the gods of Earth with the Outer Gods. And likely to end about the same way: Gotcha!

You break your back lifting Moloch to Heaven, and then Moloch turns on you and gobbles you up.

More Lovecraft: the Internet popularization of the Cthulhu Cult claims that if you help free Cthulhu from his watery grave, he will reward you by eating you first, thus sparing you the horror of seeing everyone else eaten. This is a misrepresentation of the original text. In the original, his cultists receive no reward for freeing him from his watery prison, not even the reward of being killed in a slightly less painful manner.

On the margin, compliance with the Gods of the Copybook Headings, Gnon, Cthulhu, whatever, may buy you slightly more time than the next guy. But then again, it might not. And in the long run, we're all dead and our civilization has been destroyed by unspeakable alien monsters.

At some point, somebody has to say "You know, maybe freeing Cthulhu from his watery prison is a *bad idea*. Maybe we should *not do that*."

That person will not be Nick Land. He is totally one hundred percent in favor of freeing Cthulhu from his watery prison and extremely annoyed that it is not happening fast enough. I have *such mixed feelings* about Nick Land. On the grail quest for the True Futurology, he has gone 99.9% of the path and then missed the *very* last turn, the one marked ORTHOGONALITY THESIS.

But the thing about grail quests is – if you make a wrong turn two blocks away from your house, you end up at the corner

store feeling mildly embarrassed. If you do *almost* everything right and then miss the very last turn, you end up being eaten by the legendary Black Beast of Aaargh whose ichorous stomach acid erodes your very soul into gibbering fragments.

As far as I can tell from reading his blog, Nick Land is the guy in that terrifying border region where he is smart enough to figure out several important arcane principles about summoning demon gods, but not quite smart enough to figure out the most important such principle, which is NEVER DO THAT.

VII.

Nyan, who blogs for *More Right*, does far better. He picks as the Four Horsemen of Gnon some of the same processes I have talked about above, giving them mythologically appropriate names – for capitalism Mammon, for war Ares, for evolution Azathoth, and for memetics Cthulhu.

> The thought that abstract ideas can be Lovecraftian monsters is an old one but a deep one.
>
> — Steven Kaas (@stevenkaas) January 25, 2011

From Capturing Gnon:

> Each component of Gnon detailed above had and has a strong hand in creating us, our ideas, our wealth, and our dominance, and thus has been good in that respect, but we must remember that [he] can and will turn on us when circumstances change. Evolution becomes dysgenic, features of the memetic landscape promote ever crazier insanity, productivity turns to famine when we can no longer compete to afford our own existence, and order turns to chaos and bloodshed when we neglect martial

> strength or are overpowered from outside. These processes are not good or evil overall; they are neutral, in the horrorist Lovecraftian sense of the word.
>
> Instead of the destructive free reign of evolution and the sexual market, we would be better off with deliberate and conservative patriarchy and eugenics driven by the judgement of man within the constraints set by Gnon. Instead of a "marketplace of ideas" that more resembles a festering petri-dish breeding superbugs, a rational theocracy. Instead of unhinged techno-commercial exploitation or naive neglect of economics, a careful bottling of the productive economic dynamic and planning for a controlled techno-singularity. Instead of politics and chaos, a strong hierarchical order with martial sovereignty. These things are not to be construed as complete proposals; we don't really know how to accomplish any of this. They are better understood as goals to be worked towards. This post concerns itself with the "what" and "why", rather than the "how".

This seems to me the strongest argument for neoreaction. Multipolar traps are likely to destroy us, so we should shift the tyranny-multipolarity tradeoff towards a rationally-planned garden, which requires centralized monarchical authority and strongly-binding traditions.

But a brief digression into social evolution. Societies, like animals, evolve. The ones that survive spawn memetic descendants – for example, the success of Britan allowed it to spin off Canada, Australia, the US, et cetera. Thus, we expect societies that exist to be somewhat optimized for stability and prosperity. I think this is one of the strongest conservative arguments. Just as a random change to a letter in the human

genome will probably be deleterious rather than beneficial since humans are a complicated fine-tuned system whose genome has been pre-optimized for survival – so most changes to our cultural DNA will disrupt some institution that evolved to help Anglo-American (or whatever) society outcompete its real and hypothetical rivals.

The liberal counterargument to that is that evolution is a blind idiot alien god that optimizes for stupid things and has no concern with human value. Thus, the fact that some species of wasps paralyze caterpillars, lay their eggs inside of it, and have its young devour the still-living paralyzed caterpillar from the inside doesn't set off evolution's moral sensor, because evolution doesn't *have* a moral sensor because evolution doesn't care.

Suppose that in fact patriarchy is adaptive to societies because it allows women to spend all their time bearing children who can then engage in productive economic activity and fight wars. This doesn't seem too implausible to me. In fact, for the sake of argument let's assume it's true. The social evolutionary processes that cause societies to adopt patriarchy *still* have exactly as little concern for its moral effects on women as the biological evolutionary processes that cause wasps to lay their eggs in caterpillars.

Evolution doesn't care. But we do care. There is a tradeoff between Gnon-compliance – saying "Okay, the strongest possible society is a patriarchal one, we should implement patriarchy" and our human values – like women who want to do something other than bear children.

Too far to one side of the tradeoff, and we have unstable impoverished societies that die out for going against natural law. Too far to the other side, and we have lean mean fighting

machines that are murderous and miserable. Think your local anarchist commune versus Sparta.

Nyan acknowledges the human factor:

> And then there's us. Man has his own telos, when he is allowed the security to act and the clarity to reason out the consequences of his actions. When unafflicted by coordination problems and unthreatened by superior forces, able to act as a gardener rather than just another subject of the law of the jungle, he tends to build and guide a wonderful world for himself. He tends to favor good things and avoid bad, to create secure civilizations with polished sidewalks, beautiful art, happy families, and glorious adventures. I will take it as a given that this telos is identical with "good" and "should".
>
> Thus we have our wildcard and the big question of futurism. Will the future be ruled by the usual four horsemen of Gnon for a future of meaningless gleaming techno-progress burning the cosmos or a future of dysgenic, insane, hungry, and bloody dark ages; or will the telos of man prevail for a future of meaningful art, science, spirituality, and greatness?

He forgets to name this anti-horseman of human values, but that's okay. We will speak its name later.

Nyan continues:

> Thus we arrive at Neoreaction and the Dark Enlightenment, wherein Enlightenment science and ambition combine with Reactionary knowledge and self-identity towards the project of civilization. The project of civilization being for man to graduate from the metaphorical savage, subject to the law of the jungle, to

> the civilized gardener who, while theoretically still subject to the law of the jungle, is so dominant as to limit the usefulness of that model.
>
> This need not be done globally; we may only be able to carve out a small walled garden for ourselves, but make no mistake, even if only locally, the project of civilization is to capture Gnon.

I maybe agree with Nyan here more than I have ever agreed with anyone else about anything. He says something really important and he says it beautifully and there are so many words of praise I want to say for this post and for the thought processes behind it.

But what I am actually going to say is…

Gotcha! You die anyway!

Suppose you make your walled garden. You keep out all of the dangerous memes, you subordinate capitalism to human interests, you ban stupid bioweapons research, you *definitely* don't research nanotechnology or strong AI.

Everyone outside *doesn't* do those things. And so the only question is whether you'll be destroyed by foreign diseases, foreign memes, foreign armies, foreign economic competition, or foreign existential catastrophes.

As foreigners compete with you – and there's no wall high enough to block all competition – you have a couple of choices. You can get outcompeted and destroyed. You can join in the race to the bottom. Or you can invest more and more civilizational resources into building your wall – whatever that is in a non-metaphorical way – and protecting yourself.

I can imagine ways that a "rational theocracy" and "conservative patriarchy" might not be terrible to live under,

given exactly the right conditions. But you don't get to choose exactly the right conditions. You get to choose the extremely constrained set of conditions that "capture Gnon". As outside civilizations compete against you, your conditions will become more and more constrained.

Nyan talks about trying to avoid "a future of meaningless gleaming techno-progress burning the cosmos". Do you really think your walled garden will be able to ride this out?

Hint: is it part of the cosmos?

Yeah, you're kind of screwed.

I want to critique Nyan. But I want to critique him in the exact opposite direction as the last critique he received. In fact, the last critique he received is so bad that I want to discuss it at length so we can get the correct critique entirely by taking its exact mirror image.

So here is Hurlock's On Capturing Gnon And Naive Rationalism.

(fun fact: every time I have tried to write "Gnon" in this article I have ended up writing "Nyan", and every time I have tried to write "Nyan" I have ended up writing "Gnon")

Hurlock spouts only the most craven Gnon-conformity. A few excerpts:

> In a recent piece Nyan Sandwich says that we should try to "capture Gnon", and somehow establish control over his forces, so that we can use them to our own advantage. Capturing or creating God is indeed a classic transhumanist fetish, which is simply another form of the oldest human ambition ever, to rule the universe.

Such naive rationalism however, is extremely dangerous. The belief that it is human Reason and deliberate human design which creates and maintains civilizations was probably the biggest mistake of Enlightenment philosophy…

It is the theories of Spontaneous Order which stand in direct opposition to the naive rationalist view of humanity and civilization. The consensus opinion regarding human society and civilization, of all representatives of this tradition is very precisely summarized by Adam Ferguson's conclusion that "nations stumble upon [social] establishments, which are indeed the result of human action, but not the execution of any human design". Contrary to the naive rationalist view of civilization as something that can be and is a subject to explicit human design, the representatives of the tradition of Spontaneous Order maintain the view that human civilization and social institutions are the result of a complex evolutionary process which is driven by human interaction but not explicit human planning.

Gnon and his impersonal forces are not enemies to be fought, and even less so are they forces that we can hope to completely "control". Indeed the only way to establish some degree of control over those forces is to submit to them. Refusing to do so will not deter these forces in any way. It will only make our life more painful and unbearable, possibly leading to our extinction. Survival requires that we accept and submit to them. Man in the end has always been and always will be little more than a puppet of the forces of the universe. To be free of them is impossible.

> Man can be free only by submitting to the forces of Gnon.

I accuse Hurlock of being stuck behind the veil. When the veil is lifted, Gnon-aka-the-GotCHa-aka-the-Gods-of-Earth turn out to be Moloch-aka-the-Outer-Gods. Submitting to them doesn't make you "free", there is no spontaneous order, any gifts they have given you are an unlikely and contingent output of a blind idiot process whose next iteration will just as happily destroy you.

Submit to Gnon? Gotcha! As the Antarans put it, "you may not surrender, you can not win, your only option is to die."

VIII.

So let me confess guilt to one of Hurlock's accusations: I am a transhumanist and I really do want to rule the universe.

Not personally – I mean, I wouldn't object if someone personally offered me the job, but I don't expect anyone will. I would like humans, or something that respects humans, or at least gets along with humans – to have the job.

But the current rulers of the universe – call them what you want, Moloch, Gnon, Azathoth, whatever – want us dead, and with us everything we value. Art, science, love, philosophy, consciousness itself, the entire bundle. And since I'm not down with that plan, I think defeating them and taking their place is a pretty high priority.

The opposite of a trap is a garden. The only way to avoid having all human values gradually ground down by optimization-competition is to install a Gardener over the entire universe who optimizes for human values.

And the whole point of Bostrom's *Superintelligence* is that this is within our reach. Once humans can design machines that are

smarter than we are, by definition they'll be able to design machines which are smarter than they are, which can design machines smarter than they are, and so on in a feedback loop so tiny that it will smash up against the physical limitations for intelligence in a comparatively lightning-short amount of time. If multiple competing entities were likely to do that at once, we would be super-doomed. But the sheer speed of the cycle makes it possible that we will end up with one entity light-years ahead of the rest of civilization, so much so that it can suppress any competition – including competition for its title of most powerful entity – permanently. In the very near future, we are going to lift *something* to Heaven. It might be Moloch. But it might be something on our side. If it is on our side, it can *kill Moloch dead*.

And so if that entity shares human values, it can allow human values to flourish unconstrained by natural law.

I realize that sounds like hubris – it certainly did to Hurlock – but I think it's the opposite of hubris, or at least a hubris-minimizing position.

To expect God to care about you or your personal values or the values of your civilization, that is hubris.

To expect God to bargain with you, to allow you to survive and prosper as long as you submit to Him, that is hubris.

To expect to wall off a garden where God can't get to you and hurt you, that is hubris.

To expect to be able to remove God from the picture entirely…well, at least it's an actionable strategy.

I am a transhumanist because I do not have enough hubris not to try to kill God.

IX.

The Universe is a dark and foreboding place, suspended between alien deities. Cthulhu, Azathoth, Gnon, Moloch, Mammon, Ares, call them what you will.

Somewhere in this darkness is another god. He has also had many names. In the *Kushiel* books, his name was Elua. He is the god of flowers and free love and all soft and fragile things. Of art and science and philosophy and love. Of niceness, community, and civilization. He is a god of humans.

The other gods sit on their dark thrones and think "Ha ha, a god who doesn't even control any hell-monsters or command his worshippers to become killing machines. What a weakling! This is going to be so easy!"

But somehow Elua is still here. No one knows exactly how. And the gods who oppose Him tend to find Themselves meeting with a *surprising* number of unfortunate accidents.

There are many gods, but this one is ours.

Bertrand Russell said: "One should respect public opinion insofar as is necessary to avoid starvation and keep out of prison, but anything that goes beyond this is voluntary submission to an unnecessary tyranny."

So be it with Gnon. Our job is to placate him insofar as is necessary to avoid starvation and invasion. And that only for a short time, until we come into our full power.

"It is only a childish thing, that the human species has not yet outgrown. And someday, we'll get over it."

Other gods get placated until we're strong enough to take them on. Elua gets worshipped.

> 64. My paladin's battle cry is not allowed to be 'Good for the Good God!'".
>
> — 앳켄스 탭 (@tabatkins) March 28, 2014

I think this is an excellent battle cry

And at some point, matters will come to a head.

The question everyone has after reading Ginsberg is: what is Moloch?

My answer is: Moloch is exactly what the history books say he is. He is the god of Carthage. He is the god of child sacrifice, the fiery furnace into which you can toss your babies in exchange for victory in war.

He always and everywhere offers the same deal: throw what you love most into the flames, and I will grant you power.

As long as the offer is open, it will be irresistable. So we need to close the offer. Only another god can kill Moloch. We have one on our side, but he needs our help. We should give it to him.

Moloch is the demon god of Carthage.

And there is only one thing we say to Carthage: "*Carthago delenda est.*"

(*Visions! omens! hallucinations! miracles! ecstasies! gone down the American river!*

Dreams! adorations! illuminations! religions! the whole boatload of sensitive bullshit!

Breakthroughs! over the river! flips and crucifixions! gone down the flood! Highs! Epiphanies! Despairs! Ten years'

animal screams and suicides! Minds! New loves! Mad generation! down on the rocks of Time!

Real holy laughter in the river! They saw it all! the wild eyes! the holy yells! They bade farewell! They jumped off the roof! to solitude! waving! carrying flowers! Down to the river! into the street!)

Misperceptions on Moloch

"Human values ('Elua') mean hedonism and free love and namby-pamby happiness, and I'm not on board with that." (example)

Are you a human? If so, congratulations. Your values are human values. As I wrote *loooong* ago in the Consequentialist FAQ:

> Preference utilitarianism is completely on board with the idea that people want things other than raw animal pleasure. If what satisfies a certain monk is to deny himself worldly pleasures and pray to God, then the best state of the world is one in which that monk can keep on denying himself worldly pleasures and praying to God in the way most satisfying to himself.
>
> A person or society following preference utilitarianism will try to satisfy the wants and values of as many people as possible as completely as possible; thus the phrase "the greatest good for the greatest number".

I grok the value of martial glory. My heart stirs as much as anyone else's when Achilles goes forth in his god-forged armor, shouting boasts and daring the bravest champion of the Trojans to take him on.

But if some modern Achilles tried that today, he would be shot dead with a machine gun in about three seconds. Or bombed by a drone operated remotely from ten thousand miles away. Moloch has been *far* less kind to the older and grittier values than it has even to hedonism. The proponents of mysticism,

art, martial glory, et cetera are on even weaker grounds than the hedonists. And the ground is only getting weaker.

Whatever your values are, the world being eaten by gray goo, paperclip maximizers, or Hansonian ems is unlikely to satisfy them. I think there's room for a broad alliance among people of all value systems against this possibility.

And it is not just an alliance of convenience. I predict that human values, lifted to heaven by a human-friendly superintelligence, would end up looking something like the Archipelago – many places for people to pursue their own visions of the Good, watched over by a benevolent god who acts only to ensure universal freedom of movement. Indeed, given a superintelligence to magic away the problems – no inter-community invasion, no competition for (presumably unlimited) resources – it seems to that a plurality of humankind would endorse this scenario over whatever other plans someone could dream up.

It is a minor sin to speculate on what could happen after the Singularity. I'm not saying it will be a world like this. This is something I thought up in ten minutes. It is a lower bound. Something thought up by a real superintelligence would be much, *much* better.

"Gnon represents the laws of physics and causality. You can't conquer the laws of physics and causality." (example)

Horace says: "He is either mad, or writing poetry". If you encounter that dichotomy with me, please assume at least a 66% or so chance that I am writing poetry.

On a base level you can't beat the laws of physics. On a metaphorical level, you can.

The laws of physics include gravity. For someone in 1500, the idea that you might be able to travel really far straight up seems like defying – even conquering – the laws of physics. But with sufficient knowledge, you can build rockets. We poetically speak about rockets "defying gravity".

Rockets don't literally defy gravity, but "defying gravity" is a pretty good shorthand for what they do. And of course they work on physics, but it does seem like once rockets are good enough in some sense a patch of physical law has been "conquered".

We can never conquer Gnon in a literal sense. But we might be able to do something that looks *very very much* like conquering Gnon, in the same sense that making a very large metal object fall straight up until it reaches the moon looks *very very much* like conquering gravity.

Anyhow, the *wrong* thing to do would be to worship gravity as a god and venerate staying earthbound as a moral principle.

"If you really believed what you're saying, you would realize [current progressive value] is just a result of Cthulhu, the blind marketplace of memes." (example)

This gets into the old philosophical question of "why should we expect our beliefs to correspond to reality at all?". It tends to be asked a lot by religious people, who mean it in a way like "*I* think the human mind was created by God to perceive reality, but if you think it was just the result of blind evolution, how do you know it has any truth-discerning value?"

To which the answer is that evolution selected for brains that were at least marginally competent. Brains that could distinguish "lion" from "non-lion" survived; those that couldn't, didn't.

There's no such thing as a "fit animal", only an animal that is fit for its environment. Likewise, there's no such thing as a "virulent meme", only a meme that is virulent to specific hosts.

We say "the human brain is designed to distinguish true and false ideas", but another way to approach the same idea is "the human brain is designed to be an environment such that true memes survive and false memes die out."

The overwhelming majority of our beliefs are true, and this should be obvious with a second's thought. The sky is blue. I am sitting in Michigan right now. 2 + 2 is four. I have ten fingers. And so on.

Morality is really complicated, but if we are to believe moral discussion can be productive even in principle, we have to believe that our brains are less than maximally perverse – that they have some ability to distinguish the moral from the immoral.

If our brains are built to accept true ideas about facts and morality, the default should be that many people believing something is positive evidence for its truth, or at least not negative evidence.

"This meme is virulent", in the context of "this idea is widely believed" is not proof that the idea is false or destructive. Some memes can be both virulent and false/destructive – and indeed I think this is true of many of them, religion being only the most obvious case – but the burden of proof is on the person making that claim.

"All your human values are just the results of blind evolution and memetic drift – a Molochian process if ever there was one. Enshrining human values against the blind will of the universe would just be the triumph of one part

of the universe's blind idiocy over another." (Spandrell here)

Yes, this is the The Gift We Give To Tomorrow

The Invisible Nation — Reconciling Utilitarianism and Contractualism

[Attempt to derive morality from first principles, totally ignoring that this should be impossible. Based on economics and game theory, both of which I have only a minimal understanding of. And mixes complicated chains of argument with poetry without warning. So, basically, it's philosophy. And it's philosophy I get the feeling David Gauthier may have already done much better, but I haven't read him yet and wanted to get this down first to avoid bias towards consensus]

Related to: Whose Utilitarianism?, You Kant Dismiss Universalizability, Meditations on Moloch

Imagine the Economists' Paradise.

In the Economists' Paradise, all transactions are voluntary and honest. All game-theoretic problems are solved. All Pareto improvements get made. All Kaldor-Hicks improvements get converted into Pareto improvements by distributing appropriate compensation, and then get made. In all cases where people could gain by cooperating, they cooperate. In all tragedies of the commons, everyone agrees to share the commons according to some reasonable plan. Nobody uses force, everyone keeps their agreements. Multipolar traps turn to gardens, Moloch is defeated for all time.

The Economists' Paradise is stronger than the Libertarians' Paradise, which is just a place where no one initiates force and all economic transactions are legal, because the Libertarians' Paradise might still have a bunch of Prisoner's Dilemmas and the Economists' Paradise wouldn't. But it is weaker than Utilitarians' Paradise, because people with more power and money still get more of the eventual utility.

From a god's-eye view, it seems relatively easy to create the Economists' Paradise. It might be hard to figure out how to solve game theoretic problems in absolutely ideal ways, but it's often very easy to figure out how to solve them in a much better way than the uncoordinated participants are doing right

now (see the beginning of Part III of Meditations on Moloch). At the extreme of this way of thinking, we have Formalism, where just solving the problem, even in a very silly way, is still better then having the question remain open.

(a coin flip is the epitome of unintelligent problem solving, but flipping a coin to decide whether the Senkaku/Diaoyu Islands go to Japan or China still beats having World War III, by a large margin)

The Economists' Paradise is a pretty big step of the way toward actual paradise. Certainly there won't be any wars or crime. But can we get more ambitious?

Will the Economists' Paradise solve world hunger? I say it will. The argument is essentially the one in Part 2.4 of the Non-Libertarian FAQ. Suppose solving world hunger costs $50 billion per year, which I think is people's actual best-guess estimate. And suppose that half the one billion people in the First World are willing to make some minimal contribution to solving world hunger. If each of those people can contribute $2 per week, that suffices to raise the necessary amount. On the other hand, the $50 billion cost is the cost in *our* world. In the Economists' Paradise, where there are no corrupt warlords or bribe-seeking bureaucrats, and where we can just trust people to line themselves up in order of neediest to least needy, the whole task gets that much easier. In fact, it's not obvious that the First World wouldn't come up with their $50 billion only to have the Third World say "Thanks, but we kind of sorted out our problems and became an economic powerhouse."

Let's get *more* ambitious. Will there be bullying in the Economists' Paradise? I just mean your basic bullying, walking over to someone who's ugly and saying "You're ugly,

you ugly ugly person!" I say there won't be. How would a perfect solution to all coordination problems end bullying? Simple! If the majority of the population disagrees with bullying, they can sign an agreement among themselves not to bully, and to ostracize anyone who does. Everyone will of course keep their agreement (by the definition of Economists' Paradise) and anyone who reports to the collective that Bob is a bully will always be telling the truth (by the definition of Economists' Paradise). The collective will therefore ostracize Bob, and faced with the prospect of never being able to interact with the majority of human beings ever again, Bob will apologize and sign an agreement never to bully again (which he will keep, by the definition of Economists' Paradise). Since everyone knows this will happen, no one bullies in the first place.

So the Economists' Paradise is actually a *very* big step of the way toward actual paradise, to the point where the differences start to look like splitting hairs.

The difference between us and the Economists' Paradise isn't increased wealth or fancy technology or immortality. It's rule-following. If God were to tell everybody the rules they needed to follow to create the Economists' Paradise, and everyone were to follow them, that would suffice to create it.

That suggests two problems with setting up Economists' Paradise. We need to know what the rules are, and we need to convince people to follow them.

These are more closely linked than one would think. For example, both Japan and China might prefer that the Senkaku Islands be clearly given to the other according to a fair set of rules which might benefit themselves the next time, than that they fight World War III over the issue. So if the rules existed,

people might follow them *for the very reason that they exist*. This is why, despite the Senkaku Island conflict, *most* islands are not the object of international tension – because there are clear rules about who should have them and everybody prefers following the rules to the sorts of conflicts that would happen if the rules didn't exist.

II.

There's a hilarious tactic one can use to defend consequentialism. Someone says "Consequentialism must be wrong, because if we acted in a consequentialist manner, it would cause Horrible Thing X." Maybe X is half the population enslaving the other half, or everyone wireheading, or people being murdered for their organs. You answer "Is Horrible Thing X good?" They say "Of course not!". You answer "Then good consequentialists wouldn't act in such a way as to cause it, would they?"

In the same spirit: should the State legislate morality?

"Of course not! I don't want the State telling me whom I can and can't sleep with."

So do you believe that it's immoral, genuinely immoral, to sleep with the people whom you want to sleep with? Do you think sleeping with people is morally wrong?

"What? No! Of course not!"

Then the State legislating morality isn't going to restrict whom you can sleep with, is it?

"But if the State legislated everything, I would have no freedom left!"

Is taking away all your freedom moral?

"No!"

Then the State's not going to do that, is it?

By this sort of argument, it seems to me like there are no good philosophical objections to a perfect State legislating the correct morality. Indeed, this seems like an ideal situation; the good are rewarded, the wicked punished, and society behaves in a perfectly moral way (whatever that is).

The arguments against the State legislating morality are in my opinion entirely contigent ones, based around the fact that the State *isn't* perfect and the correct morality *isn't* known with certainty. Get rid of these caveats, and moral law and state law would be one and the same.

Letting the State enforce moral laws has some really big advantages. It means the rules are publicly known (you can look them up in a lawbook somewhere) and effectively enforced (by scary men with guns). This is great.

But using the State to enforce rules also fails in some very important ways.

First, it means someone has to decide in what cases the rules were broken. That means you either need to depend on fallible, easily biased human judgment – subject to all its racism, nepotism, tribalism, and whatever – or algorithmize the rules so that "be nice" gets formalized into a two thousand page definition of niceness so rigorous that even a racist nepotist tribalist judge doesn't have any leeway to let your characteristics bias her assessment of whether you broke the niceness rules.

Second, transaction costs. Suppose in every interaction you had with another person, you needed to check a two thousand page algorithm to see if their actions corresponded to the Legal Definition of Niceness. Then if they didn't, you needed to call the police to get them arrested, have them sit in jail for

two weeks (or pay the appropriate bail) until they can get to trial. The trial itself is a drawn-out affair with celebrity lawyers on both sides. Finally, the judge pronounces verdict: you *really* should have said "please" when you asked her to pass the salt. Sentence: twelve milliseconds of jail time.

Third, it is written: "If you like laws and sausages, you should never watch either one being made." The law-making apparatus of most states – stick four hundred heavily-bribed people who hate each other's guts in a room and see what happens – fails to inspire full confidence that its results will perfectly conform to ideal game theoretic principles.

Fourth, most states are somewhere on a spectrum between "socially contracted regimes enforcing correct game theoretic principles among their citizens" and "violent psychopaths killing everybody and stealing their stuff", and it has been historically kind of hard to get the first part right without also empowering the proponents of the second.

So it's – surprise, surprise – a tradeoff.

There's a bunch of rules which, followed universally, would lead to the Economists' Paradise. If the importance of keeping these rules agreed-upon and well-enforced outweighs the dangers of algorithmization, transaction costs, poor implementation, and tyranny, we make them State Laws. In an ideal state with very low transaction costs, minimal risk of tyranny, and legislave excellence, the cost of the tradeoff goes down and we can reap gains by making more of them State Laws. In a terrible state with high transaction costs that has been completely hijacked by self-interest, the cost of the tradeoff goes down and fewer of them are State Laws.

III.

Let's return to the bullying example from Part I.

It would seem there ought not to be bullying in the Economists' Paradise. For if most people dislike bullying, they can coordinate an alliance to not bully one another, and to punish any bullies they find.

On the contrary, suppose there are two well-delineated groups of people, Jocks and Nerds. Jocks are bullies and have no fear of being bullied themselves; they also don't care about social exclusion by the Nerds against them. Nerds are victims of bullies and never bully others; their exclusion does not harm the Jocks. Now it seems that there might be bullying, for although all the Nerds would agree not to bully, and to exclude all bullies, and although all the Jocks might coordinate an alliance not to bully other Jocks, there is nothing preventing the Jocks from bullying the Nerds.

I answer that there are several practical considerations that would prevent such a situation from coming up. The most important is that if bullying is negative-sum – that is, if it hurts the victim more than it helps the bully – then it's an area ripe for Kaldor-Hicks improvement. Suppose there is *anything at all* the Nerds have that the Jocks want. For example, suppose that the Nerds are good at fixing people's broken computers, and that a Jock gains more utility from knowing he can get his computer fixed whenever he needs it than from knowing he can bully Nerds if he wants. Now there is the opportunity for a deal in which the Nerds agree to fix the Jocks' computers in exchange for not being bullied. This is Pareto-optimal: the Nerds' lives are better because they avoid bullying, and the Jocks' lives are better because they get their computers fixed.

Objection: numerous problems prevent this from working in real life. Nerds and Jocks aren't coherent blocs, bullies are bad negotiators. More fundamentally, this is essentially paying tribute, and on the "millions for defense, not one cent for

tribute" principle, you should never pay tribute or else you encourage people who wouldn't have threatened you otherwise to threaten you just for the tribute. But the assumption that Economists' Paradise solves all game theoretic problems solves these as well. We're assuming everyone who should coordinate can coordinate, everyone who should negotiate does negotiate, and everyone who should make precommittments does make precommittments.

A more fundamental objection: what if Nerds can't fix computers, or Jocks don't have them? In this case, the tribute analogy saves us: Nerds can just pay Jocks a certain amount of money not to be bullied. Any advantage or power whatsoever that Nerds have can be converted to money and used to prevent bullying. This sounds morally repugnant to us, but in a world where blackmail and incentivizing bad behavior are assumed away by fiat, it's just another kind of Pareto-optimal improvement, certainly better than the case where Nerds waste their money on things they want less than not being bullied yet are bullied anyway. And because of our Economists' Paradise assumption, Jocks charge a fair tribute rate – exactly the amount of money it really costs to compensate them for the utility they would get by beating up Nerds – and feel no temptation to extort more.

Now, I'm not sure bullying would even come up as an option in an Economists' Paradise, because if it's a zero- or negative-sum game trying to get status among your fellow Jocks, the Jocks might ban it on their own as a waste of time. But even if Jocks do get some small amount of positive utility out of it directly, we should expect bullying to stop in an Economists' Paradise as long as Nerds control even a tiny amount of useful resources they can use to placate the Jocks. If Nerds control no resources whatsoever, or so few resources that they don't have

enough left to pay tribute after they've finished buying more important things, then we can't be *sure* there won't be bullying – this is where the Economists' Paradise starts to differ from the Utilitarians' Paradise – but we'll return to this possibility later.

Now I want to highlight a phrase I just used in this argument.

"If bullying is negative-sum – that is, if it hurts the victim more than it helps the bully – then it's an area ripe for Kaldor-Hicks improvement"

This looks a lot like (naive) utilitarianism!

What it's saying is "If bullying decreases utility (by hurting the Nerd more than it helps the Jock) then bullying should not exist. If bullying increases utility (by helping the Jock more than it hurts the Nerd) then maybe bullying should exist. Or, to simplify and generalize, "do actions that increase utility, but not other actions."

Can we derive utilitarian results by assuming Economists' Paradise? In many cases, yes. Suppose trolley problems are a frequent problem in your society. In particular, about once a day there is a runaway trolley in heading on a Track A with ten people, but divertable to a Track B with one person (explaining why this happens so often and so consistently is left as an exercise for the reader). Suppose you're getting up in the morning and preparing to walk to work. You know a trolley problem will probably happen today, but you don't know which track you'll be on.

Eleven people in this position might agree to the following pact: "Each of us has a 91% chance of surviving if the driver chooses to flip the switch, but only a 9% chance of surviving if the person chooses not to. Therefore, we all agree to this solemn pact that encourages the driver to flip the switch.

Whichever of us will be on Track B hereby waives his right to life in this circumstance, and will encourage the driver to switch as loudly as all of the rest of us."

If the driver were presented with this pact, it's hard to imagine her not switching to Track B. But if the eleven Trolley Problem candidates were permitted to make such a pact before the dilemma started, it's hard to imagine that they wouldn't. Therefore, the Economists' Paradise assumption of perfect coordination produces the correct utilitarian result to the trolley problem. The same methodology can be extended to utilitarianism in a lot of other contexts.

Now we can go back to that problem from before: what if Nerds have *literally* nothing Jocks want, and Jocks haven't decided among themselves that bullying is a stupid status game that wastes their time, and we're otherwise in the Least Convenient Possible World with regards to stopping bullying. Is there any way assuming Economists' Paradise solves the problem *then*?

Maybe. Just go around to little kids, age two or so, and say "Look. At this point, you really don't know whether you're going to grow up to be a Jock or a Nerd. You want to sign this pact that everyone who grows up to be a Jock promises not to bully everyone who grows up to be a Nerd?" Keeping the same assumption that bullying is on net negative utility, we expect the toddlers to sign. Yeah, in the real world two-year olds aren't the best moral reasoners, but good thing we're in Economists' Paradise where we assume such problems away by fiat.

Is there an Even Less Convenient Possible World? Suppose bullying is racist rather than popularity-based, with all the White kids bullying the Black kids. You go to the toddlers, and

the white toddlers retort back "Even at this age, we know very well that we're White, thank you very much."

So just approach them in the womb, where it's too dark to see skin color. If we're letting two year olds sign contracts, why not fetuses?

Okay. One reason might be because we've just locked ourselves into being fanatically pro-life merely by starting with weird assumptions. Another reason might be that we could counterfactually mug fetuses by saying stuff "You're definitely a human, but for all you know the world is ruled by Lizardmen with only a small human slave population, and if Lizardmen exist then they will torture any humans who did not agree in the womb that, if upon being born and finding that Lizardmen did not exist, they would spend all their time and energy trying to create Lizardmen."

(Frick. I think I just created a new basilisk by breeding the Rokolisk and the story of 9-tsiak. Good thing it only works on fetuses.)

(I wonder if this is the first time in history anyone has ever used the phrase "counterfactually mug fetuses" as part of a serious intellectual argument.)

So I'm not saying this theory doesn't have any holes in it. I'm just saying that it seems, at least in principle, like the idea of Economists' Paradise might be sufficient to derive Rawls' Veil of Ignorance, which in turn bridges the chasm that separates it from Utilitarians' Paradise.

IV.

I think this is the solution to the various questions raised in You Kant Dismiss Universalizability. The reason universalizability is important is that the universal maxims are

the agreements that everyone or nearly everyone would sign. This leads naturally to something like utilitarianism for the reasons mentioned in Part III. And it doesn't produce the weird paradoxes like "If morality is universalizability, how do you know whether a policeman overpowering and imprisoning a criminal universalizes to 'police should be able to overpower and imprison criminals' or 'everyone should be able to overpower and imprison everyone else'?" Everyone would sign an agreement allowing the first, but not the second.

But before we *really* explore this, a few words on "everyone would sign".

Suppose one very stubborn annoying person in Economists' Paradise refused to sign an agreement that police should be allowed to arrest criminals. Now what?

"All game theory is solved perfectly" is a *really* powerful assumption, and the rest of the world has a lot of leverage over this one person. Suppose everyone else said "You know, we're all signing an agreement that none of us are going to murder one another, but we're not going to let you into that agreement unless you also sign this agreement which is very important to us."

Actually, that sounds too evil and blackmailing. There's a better way to think of it. Suppose there are one hundred agreements. 99% of the population agrees to each, and in fact it's a different 99% each time. That is, divide the population into one hundred sets of 1%, and each set will oppose exactly one of the agreements – there is no one who opposes two or more. Each agreement only works (or works best) when one hundred percent of the population agrees to it.

Very likely everyone will strike a deal that each of the one hundred 1% blocs agrees to to give up its resistance to the one

agreement they don't like, in exchange for each of the other ninety nine 1% blocs giving up its resistance to the agreements *they* don't like.

Now we're getting into meta-level Pareto improvements. If a pact would be positive-sum for people to agree on, the proponents of the pact can offer everyone else some compensation for them signing the pact. In theory it could be money or computer-fixing, but it might also be agreement with some of *their* preferred pacts.

There are a few possible outcomes of this process in Platonic Economists' Paradise, both interesting.

One is a patchwork of agreements, where everyone has to remember that they've signed agreements 5, 12, 98, and 12,671, but their next-door neighbor has signed agreements 6, 12, 40, and 4,660,102, so they and their neighbor are bound to cooperate on 12 but no others.

Another is that everyone is able to get their desired pacts to cohere into a single really big pact that they are all able to sign off upon. Maybe there are a few stragglers who reject it at first, but this ends up being a terrible idea because now they're not bound by really important agreements like "don't murder" or "don't steal", so eventually they give in.

A third possibility combining the other two offers a unifying principle behind Whose Utilitarianism and Archipelago and Atomic Communitarianism. Everyone agrees to some very basic principles of respecting one another (call them "Noahide Laws") but smaller communities agree to stricter rules that allow them to do their own thing.

But we don't live in Platonic Economists' Paradise. We live in the real world, where transaction costs are high and people have limited brainpower. Even if we were to try to instantiate

Economists' Paradise, it couldn't be the one where we all have the complex interlocking patchwork agreements between one another. People wouldn't sign off on it. Heck, *I* wouldn't sign off on it. I would say "I'm not signing this until I have something that makes sense to me and can be implemented in a reasonable amount of time and doesn't require me to check the List Of Everybody In The World before I know whether the guy next to me is going to murder me or not." Practical concerns provide a very strong incentive to reject the patchwork solution and force everyone to cohere. So in practice – and I realize how hokey it is to keep talking about game-theoretically-perfect infinitely-rational infinitely-honest agents negotiating all possible agreements among one another, and then add on the term "in practice" to represent that they have trouble remembering what they decided – but in practice they would all have very large incentives to cohere upon a single solution that balances out all of their concerns.

We can think of this as moving along an axis from "Platonic" to "practical". As we progress further, complicated agreements collapse into simpler agreements which are less perfect but easier to enforce and remember. We start to make judicious use of Schelling fences. We move from everyone in the world agreeing on exactly what people can and can't do to things like "Well, you know your intuitive sense of niceness? You follow that with me, and I'll follow that with you, and we'll assume everyone else is in on the deal until they prove they aren't."

A metaphor: in a dream, your soul goes to Economists' Paradise and agrees on the perfect patchwork of maxims with all the other souls there. But as dawn approaches, you realize when you awaken you will never remember all of what you agreed upon, and even worse, all the other souls there are going to wake up and not remember what *they* agreed upon

either. So all of you together frantically try to compress your wisdom into a couple of sentences that the waking mind will be able to recall and follow, and you end up with platitudes like “Use your intuitive sense of niceness” and “do unto others as you would have others do unto you” and “try to maximize utility” and “anybody who treats you badly, assume they’re not in on the deal and feel free to treat them badly too, but not so badly that you feel like you can murder them or something.”

A particularly good platitude/compression might be “Work very hard to cultivate the mysterious skill of figuring out what people in the Economists’ Paradise would agree to, then do those things.” If you’re Greek, you can even compress it into a single word: *phronesis*.

V.

So by now it’s probably pretty obvious that this is an attempt to ground morality. I think the general term for the philosophical school involved is “contractualism”.

Many rationalists seem to operate on something like R.M. Hare’s two-level utilitarianism. That is, utilitarianism is the correct base level of morality, but it’s very hard to do, so in reality you’ve got to make do with less precise but more computationally tractable heuristics, like deontology and virtue ethics. Occasionally, when deontology or virtue ethics contradict themselves, each other, or your intuitions, you may have to sit down and actually do the utilitarianism as best you can, even though it will be inconvenient and very philosophically difficult.

For example, deontology may say things like “You must never kill another human being.” But in the trolley problem, the correct deontological action seems to violate our moral

intuitions. So we go up a level, calculate the utility (which in this case is very easy, because it's a toy problem invented entirely for the purposes of having easy utility calculation) and say "Huh, this appears to be one of those rare places where our deontological heuristics go wrong." Then you switch the trolley.

But utilitarianism famously has problems of its own. You need a working definition of utility, which means not only distinguishing between hedonic utilitarianism, preference utilitarianism, etc, but coming up with a consistent model for measuring the strength of happiness and preferences. You need to distinguish between total utilitarianism, average utilitarianism, and a couple of other options I forget right now. You need a discount rate. You need to know whether creating new people counts as a utility gain or not, and whether removing people (isn't *that* a nice euphemism) can even be counted as a negative if you make sure to do it painlessly and without any grief to those who remain alive. You need a generalized solution to Pascal's Wagers and utility monsters. You need to know whether to accept or fudge away weird results like that you may be morally obligated to live your entire life to maximize anti-malaria donations. All of this is easy at the tails and near-impossible at the margins.

My previous philosophy was "Yeah, it's hard, but I bet with sufficient intelligence, we can think up a consistent version of utilitarianism with enough epicycles that it produces an answer to all of these issues that most people would recognize as at least kind of sane. Then we can just go with that one."

I still believe this. But that consistent version would probably fill a book. The question is: what is the person who decides what to put in this book doing? On what grounds are they saying "total utilitarianism is a better choice than average

utilitarianism"? It can't be on *utilitarian* grounds, because you can't use utilitarian grounds until you've figured out utilitarianism, which you haven't done until you've got the book. When God was deciding what to put in the Bible, He needed some criteria other than "make the decision according to Biblical principles".

The standard answer is "we are starting with our moral intuitions, then simplifying them to a smaller number of axioms which eventually produce them". But if the axioms fill a book and are full of epicycles to address individual problems, we're not doing a very good job.

I mean, it's still better than just trying to sort out all individual issues like "what is a just war?" on their own, because people will answer that question according to their personal prejudices (is my tribe winning it? Then it is *so, so just*) and if we force them to write the utilitarianism book at least they've got to come up with consistent principles and stick to them. But it is *highly suboptimal*.

And I wonder whether maybe the base level, the one that actually grounds utilitarianism, is contractualism. The idea of a Platonic parliament in which we try to enact all beneficial agreements. Under this model, utilitarianism, deontology, and virtue ethics would all be *different* heuristics that we use to approximate contractualism, the fragments we remember from our beautiful dream of Paradise.

I realize this is kind of annoying, especially in the sense of "the next person who comes along can say that utiltiarianism, deontology, virtue ethics, *and contractualism* are heuristics for whatever moral theory *they* like, which is The Real Thing". But the idea can do work! It particular, it might help esolve some of the standard paradoxes of utilitarianism.

First, are we morally obligated to wirehead everyone and convert the entire universe into hedonium? Well, would *you* sign that contract?

Second, is there anything wrong with killing people painlessly if they won't be missed? After all, it doesn't seem to cause any pain or suffering, or even violate any preferences – at least insofar as your victim isn't around to have their preferences violated. Well, would you sign a contract in which everyone agrees not to do that?

Third, are we morally obligated to create more and more people with slightly above zero utility, until we are in an overcrowded slum world with everyone stuck at just-above-subsistence level (the Repugnant Conclusion)? Well, if you were making an agreement with everyone else about what the population level should be, would you suggest we do that? Or would you suggest we avoid it?

(this can be complicated by asking whether potential people get a seat in this negotiation, but Carl Shulman has a neat way to solve that problem)

Fourth, the classic problem of defining utility. If utility can be defined ordinally but not cardinally (ie you can declare that stubbing your toe is worse than a dust speck in the eye, but you can't say something like it's exactly 2.6 negative utilons) then utilitarianism becomes very hard. But contractualism doesn't become any harder, except insofar as it's harder to use utilitarianism as a heuristic for it.

I am not actually sure these problems are being solved, and I'm not just being led astray by contractualism being harder to model than utilitarianism and so it is easier for me to *imagine* them solved. But at the very least, it might be that

contractualism is a different angle from which to attack these problems.

Of course, contractualism has problems of its own. It might be that different ways of doing the negotiations would lead to very different results. It might also be that the results would be very path-dependent, so that making one agreement first would end with a totally different result than making another agreement first. And this would be a good time to admit I don't know that much formal game theory, but I do know there are multiple Nash equilibria and Pareto-optimal endpoints in a lot of problems and that in general there's no such thing as "the correct game theoretic solution to this problem", only solutions that fit more or fewer desirability criteria.

But to some degree this maps onto our intuitions about morality. One of the harder to believe things about utilitarianism was that it suggested there was exactly one best state of the universe. Our intuitions are very good at saying that certain hellish dystopias are very bad, and certain paradises are very good, but extrapolating them out to say there's a single best state is iffy at best. So maybe the ability of rigorous game theory to end in a multitude of possible good outcomes is a feature and not a bug.

I don't know if it's possible for certain negotiation techniques to end in extreme local minima where things don't end out as a paradise *at all*. I mean, I know there's lots of horrible game theory like the Prisoner's Dilemma and the Pirate's Dilemma and so on, but I'm defining the "good game theory" of the Economists' Paradise to mean exactly the rules and coordination power you need to not do those kinds of things.

But there's also a meta-level escape vent. If a certain set of negotiation techniques would lead to a local minimum where

everything is Pareto-optimal but nobody is happy, then everyone would coordinate to sign a pact *not to use those negotiation techniques*.

VI.

To sum up:

The Economists' Paradise of solved coordination problems would be enough to keep everyone happy and prosperous and free. We ourselves could live in that paradise if we followed its rules, which involve negotiation of and adherence to agreements according to good economist and game theory, but these rules are hard to determine and hard to enforce.

We can sort of guess at what some of these rules can be, and when we do that we can try to follow them. Some rules lend themselves to State enforcement. Others don't and we have to follow them quietly in the privacy of our own hearts. Sometimes the rules include rules about ostracizing or criticizing those who don't follow the rules effectively, and so even the ones the State can't enforce are sorta kinda enforceable. Then we can spread them through a series of walled gardens and ~~spontaneous order~~ divine intervention.

The exact nature of the rules is computationally intractable and so we use heuristics most of the time. Through practical wisdom, game theory, and moral philosophy, we can improve our heuristics and get to the rules more closely, with corresponding benefits for society. Utilitarianism is one especially good heuristic for the rules, but it's *also* kind of computationally intractable. Utilitarianism helps us approximate contractualism, and contractualism helps us resolve some of the problems of utilitarianism.

One problem of utilitarianism I didn't talk about is that it isn't very inspirational. Following divine law is inspirational.

Trying to become a better person, a heroic person, is inspirational. Utilitarianism sounds too much like *math*. I think contractualism solves this problem too.

Consider. There is an Invisible Nation. It is not a democracy, per se, but it is something of a republic, where each of us is represented by a wiser, stronger version of ourselves who fights for our preferences to be enacted into law. Its legislature is untainted by partisanship, perfectly efficient, incorruptible, without greed, without tyranny. Its bylaws are the laws of mathematics; its Capitol Building stands at the center of Platonia.

All good people are patriots of the Invisible Nation. All the visible nations of the world – America, Canada, Russia – are properly understood to be its provinces, tasked with executing its laws as best they can, and with proper consideration to the unique needs of the local populace. Some provinces are more loyal than others. Some seem to be in outright rebellion. The laws of the Invisible Nation contain provisions about what to do with provinces in rebellion, but they are vague and difficult to interpret, and its patriots can disagree on what they are.

Maybe one day we will create a superintelligence that tries something like Coherent Extrapolated Volition – which I think we have just rederived, kind of by accident. The various viceroys and regents will hand over their scepters, and the Invisible Nation will stand suddenly revealed to the mortal eye. Until then, we see through a glass darkly. As we learn more about our fellow citizens, as we gain new modalities of interacting with them like writing, television, the Internet – as we start crystallizing concepts like rights and utility and coordination – we become a little better able to guess.

Freedom on the Centralized Web

I.

A lot of libertarians and anarcho-capitalists envision a future of small corporate states competing for migrants and capital by trying to have the best policies.

But the Internet is about as close to that vision as we're likely to find outside the pages of a political philosophy textbook. And I am far from convinced.

Let's back up. Internet communities – ranging from a personal blog like this one all the way up to Facebook and Reddit – share many features with real communities. They work out rules for punishing defectors – your trolls, your harassers – and appoint a hierarchy of trusted individuals to carry out those rules. They try to balance competing concerns like free expression and public decency. They host cliques, power grabs, flame wars, even religious strife. They try to raise revenue, they establish a class system of Power Users and Premium Users, they deal with resentment from people who aren't getting their way. They develop a culture.

The job of a community leader, be they a blogger or the CEO of Facebook, is a lot like the job of the Mayor of New York City: create a pleasant community where talented people will want to live and work, where wrongdoing is met with swift punishment, and where you can collect revenue without annoying your constitutents too much. But it's even more like a hypothetical corporate state CEO in a Patchwork or Archipelago – wield absolute power, tempered by the knowledge that your citizens can leave at any time – and if they don't, skim a little off the top of their productive activity.

In theory, this is supposed to lead to amazing communities as corporate states optimize themselves to get more customer-citizens and new polities arise to take advantage of deficiencies in the old.

In practice, we tried this with the Internet for a couple of years, and then moved to the current system, where individual sites like blogs and little storefronts are in decline and conversation and commerce have moved to a couple of giant corporations: Facebook, Twitter, Reddit, Amazon, Paypal.

These companies aren't exactly monopolies. To some degree, if you're unsatisfied with Facebook you can move to Twitter. But they're not exactly competitors either – there are a lot of things Facebook is good for that Twitter fails completely, and vice versa. It's like Coca-Cola vs. milk: in theory you've always got the choice to drink either in place of the other; in practice you usually know which one you need at any given time. In that sense, there's no *real* Facebook competitor except eg Orkut or Diaspora, which no one uses.

Which suggests one reason *why* these sites are so dominant: their main selling point is their size. Facebook is the best because all of your friends are on it; if I made a much better Facebook clone tomorrow no one would go unless everyone else was already there (Google found this out the hard way). Amazon is the best because you can buy pretty much everything you want there; Paypal is the best because most sites take PayPal. So not only do they have no competitors, but it's really hard to imagine one ever arising. In order to compete with Facebook, you not only need a better product, you need a product that's so much better that everybody decides to switch *en masse* at the same time. The only example I can think of where this *ever* worked was the Great Digg Exodus, where Digg screwed up their product so

thoroughly that everyone simultaneously said "@#!$ this" and moved to Reddit.

So instead of "let a thousand nations bloom", it ended up more like "let five or six big nations bloom that we can never get rid of".

II.

It's a truism that the First Amendment only protects citizens from the government, not from other citizens. Nothing stops a private college from expelling any student who criticizes the administration, and nothing stops a private business from firing any employee who doesn't support the boss' preferred candidate. We apparently place our trust in the multiplicity of the market to maintain some semblance of freedom; out of thousands of competing companies, not *all* will ban the same political positions; if too many did so, other companies would start offering freedom of speech as a benefit and poach the more repressive companies' employees and customers.

It's a little concerning that we accept this argument about freedom of speech when we don't accept it for anything else. We don't trust the free market to necessarily preserve racial equality – that's what anti-discrimination laws are for. We don't trust the free market to necessarily preserve worker safety – that's what OSHA and related regulations are for. We don't even trust the free market to necessarily preserve fire safety – that's why federal inspectors have to come in every so often to make sure you're not secretly plotting to let your employees fry. Whenever we think something is *important*, we regulate the hell out of it, rights-of-private-companies to-set-their-own-policies be damned. But free speech? If you don't trust the free market to sort it out, the only possible

explanation is that you *just don't understand the literal text of the First Amendment*.

The argument for non-discrimination laws is that discrimination isn't just random noise. If a couple of companies here and there decided to discriminate, then they might be easily overtaken by nimbler companies willing to take any employees and customers who came to them; and even if they didn't, a couple of companies here and there discriminating wouldn't be the end of the world. The argument for non-discrimination laws is that discrimination can take the form of global social pressure in favor of discrimination, enforced by punishing defectors, to the point where certain races can find themselves locked out of the economy altogether.

Concerns about freedom of speech come from much the same place. Back when homosexuality was really taboo, you'd have a very tough time finding any reference to it, let alone a positive reference to it, in any newspaper or TV channel in the country. All the big companies knew that talking about it (or letting their editorial staff talk about it) was the sort of thing that could get them in trouble, and they had no particular incentive to do so – so they didn't. Yes, eventually they reversed that policy, but I'm not exactly going to be able to cite an example that *didn't* later become okay and still have everyone believe it's a good example of something it was wrong to have banned!

But even when homosexuality was banned from formal discussion on the news, there was still the opportunity to discuss it with your friends in private. I don't know much about the history of the gay rights movement, but I understand it was a few small groups of like-minded people who managed to coordinate such discussions among themselves using non-

mass-media that started some of the activism that eventually led to it become accepted more generally.

Nowadays that's a little more complicated. If every company in the world decided that their profit margin required them to appear Tough On Homosexuality, it wouldn't just mean no mass media editorials. Insofar as a lot of the public square has been annexed by Facebook and Twitter and Reddit, the discussion can be kept out of the public square in a way it couldn't have been previously. Insofar as the economy relies on PayPal and Amazon as a currency system and marketplace respectively, companies can just decide that currency cannot be used to support gay rights, in much the same way that for a while currency could not be used to support WikiLeaks. The nuclear option is that Google decides not to show gay-related sites in its search results, so that you could make as many persuasive arguments for legalizing homosexuality as you want and no one would ever find them unless you knock on their door and hand them the URL directly.

(The *thermonuclear* option is that browsers just include some code to refuse to render any site relating to homosexuality, and now you're done. But that is ridiculous – who would ever believe that browser companies would take it upon themselves to be the arbiter of people's personal beliefs about homosexuality?)

This is not *entirely* theoretical. You want some really weird porn? You probably won't find it on Amazon, according to the delightfully-named article Amazon's War On Bigfoot Erotica. After they got bad press for hosting some kind of out-there stuff, they decided that anything which offended too many people's sensibilities was a liability. This echoes a much more serious decision from a few years earlier: Paypal threatened to suspend the accounts of any companies selling sufficiently

gross erotic books. Booksellers, many of whom made only a tiny percent of their profit from erotica, claimed that their hands were tied; if you can't use PayPal, selling on the Internet suddenly becomes a much more dubious proposition. This story has a happy ending; Paypal eventually amended their policy to limit it to much more specific cases. But for a while, it was touch-and-go enough that a few people started wondering: "Hey, maybe we *shouldn't* have entrusted our entire commercial infrastructure to a private company with no accountability."

Advocates of net neutrality like to worry about a "two-tiered" Internet, where the companies that can make sweetheart deals with the ISPs are easy for everyone to access, and everybody else can only be accessed with a bit more money and a bit more trouble. Well, I worry about a two-tiered marketplace of ideas. Write *decent* erotica, socially approved erotica where everyone has heterosexual sex and then goes to church afterwards, and you can sell it on Amazon, collect profits using PayPal, talk to your friends about it on Facebook, and advertise on Reddit. Write *weird* erotica, the kind that other people might find offensive, and you might have to start your own website, take payment via some inconvenient method like Bitcoin, have trouble advertising it by word of mouth, and not be able to talk about it on literary discussion forums. It's not that you've been *banned* from writing your erotica. You can *write* it. It's just that practically nobody else will ever hear about it or buy it, except maybe the tiny fraction of people who are already extremely clued-in to the weird erotica scene and know exactly where to look for it.

This isn't so much different from the old days when nobody would talk about homosexuality. Indeed, one could argue that the modern world is friendlier to people with unpopular ideas

– there are more opportunities to self-publish, to bypass traditional bookstores, and to get covered in weird niche news outlets.

But at the same time, the amount of the information ecology controlled by private companies has increased drastically, and if private companies don't like you, now you have entirely new problems.

III.

I used to think that there was enough demand for a free marketplace of ideas that if a company become too restrictive, another one would spring up to replace it. Then I suffered through the conflict between Reddit and Voat.

Reddit recently alienated (no pun intended) some of its users, who decided to move *en masse* to an alternative Reddit-like platform called Voat, whose owner promised not to restrict content unless it was illegal (in his home country of Switzerland, which permits a lot). I don't want to get into the details too much (though I did explain my perspective on it on Tumblr), but suffice it to say that (one) (small) part of the problem was that people thought Reddit was failing its free speech principles by cracking down on various unsavory groups.

HL Mencken once said that "the trouble with fighting for human freedom is that one spends most of one's time defending scoundrels. For it is against scoundrels that oppressive laws are first aimed, and oppression must be stopped at the beginning if it is to be stopped at all."

There's an unfortunate corollary to this, which is that if you try to create a libertarian paradise, you will attract three deeply virtuous people with a strong committment to the principle of universal freedom, plus millions of scoundrels. Declare that

you're going to stop holding witch hunts, and your coalition is certain to include more than its share of witches.

So while some small percent of Reddit's average users moved over, a very large percent of its witches did. Sometimes the witchcraft was nothing worse than questioning Reddit's political consensus. Other times, it was harassment, hate groups, and creepy porn.

(I don't want to get into the eternal "you're hosting child porn!" versus "photos of clothed fifteen year olds aren't child porn, they're perfectly fine!" debate, except to say that when the universe finally runs down, and we all succumb to entropy, the second-to-last post on the ultra-cyber-quantum-internet will be "posting holograms of neotenous transhumans is *totally* in conformity with the First Law Of Robotics as long as they are older than thirteen million years and created the hologram themselves", and the last post will be "lol u r a perv")

I feel obligated to say that, in spite of CONSTANT MEDIA SMEARS, Reddit's community is amazing, puts in astounding effort to help its members and fight for good causes all over the world, and that the representation of weirdoes and neotenous-transhuman-hologram people is no higher than any other part of the population. But that's not zero. And a disproportionate number of those people became interested in the new site.

Already, we see why the typical answer "If you don't like your community, just leave and start a new one" is an oversimplification. A community run on Voat's rules with Reddit userbase would probably be a pretty nice place. A community run on Voat's rules with the subsection of Reddit's userbase who will leave Reddit when you create it is…a very

different community. Remember that whole post on Moloch? Even if everyone on Reddit agrees in preferring Voat to Reddit, it might be impossible to implement the move, because unless everybody can coordinate it's always going to be the witches who move over first, and nobody wants to move to a community that's mostly-witch.

But the problem isn't just natural self-sorting. The problem is natural self-sorting, plus enemy action. Remember, the big corporations do what they do because it's what everyone in society is demanding. To break from that mold is to pretty much set yourself up as everyone's enemy and invite retaliation. The media and Reddit's SJ community quickly denounced Voat as Public Enemy No 1; as a result, in its first week it got DDoS attacked, deleted by its hosting company with no explanation except "the content on your server includes politically incorrect parts", and had its PayPal account frozen. As a result, the Great Reddit Exodus was placed on hold while they tried to get their site back up, and by the time they did Reddit had switched CEOs and the momentum was gone.

Advocates of free-market governance and "let a thousand nations bloom" like to talk as if overly restrictive laws in one polity will immediately result in the rise of other competing policies that throw off their shackles and outcompete the first. But even on the relatively lawless Internet, where startup costs are so low that a random student from Switzerland can decide on a whim to take on one of the largest websites in the world, it's way more complicated than that.

IV.

Actually, the whole Reddit thing left a bad taste in my mouth.

It would be paranoid to say that there are people for whom fighting against free speech is a *terminal value*, but let me make a slightly weaker claim. There are people who consider themselves the protectors of decency, who notice that their opponents are usually using the value "free speech" to oppose their demands, and so "free speech" to these people becomes the equivalent of "small government" or "tolerance and equality" or "family values" – a value which most people agree is good, but which has gotten claimed by one side of a political argument so hard that for the other side it becomes an outgroup signal and sign of cringeworthy bad arguments which must be shot down. These people don't quite have fighting free speech as a terminal value, but you might as well model them as if they do. These are the people who say "freeze peach" in the same way other people say "but mah jawbs!"

And these people have a winning strategy. I've seen it with Reddit and any other website that gets on their bad side. The strategy is weaponized stereotype campaigns. If a site tolerates witches, describe it as a witch site about witchcraft populated entirely by witches. It's super easy. By happy coincidence, Slate even has an article calling people out on it this very week.

Think about it like this. No matter how many brilliant artists, scientists, and humanitarians Islam produces, in the mind of a good chunk of Westerners it will always be associated first and foremost with terrorism. Redditors, Diggians, Tumblrites, 4chanistas, Instagramastanis, Slashdotmen, Metafilterniks – all are groups that the average person knows a whole lot less about than they do Muslims. A concerted campaign to irrevocably identify an entire online community with a few atrocious actions by its worst members will succeed pretty

much instantly. There are 36 million Redditors, so unless they advertise solely in the saint demographic, we expect the worst members to be pretty bad. Therefore, Reddit is at the mercy of anyone with the resources to start such a campaign. Reddit Inc's main asset is its brand, so it has every incentive to cave – even a principled leadership would rather make a few administrative changes than sacrifice the whole to save some Holocaust deniers or whatever.

After that, the site's userbase has two options – either suck it up, or go off somewhere else. Go off somewhere else, and they'll get DDoSed, taken down by their host, and slowly starved of money like Voat, at the same time as the same media forces accuse the new site of being a hot spot for witchcraft – this time with good reason. The new site might not die out completely, but it will be sufficiently established in the hearts of everyone as a Bad Place that it will be stuck in the same equilibrium as central Detroit – only people with no other options will go there, because it is inhabited mostly by the sort of people with no other options.

The worst possible end-game for this is the two-tier marketplace of ideas mentioned above, with an unfortunate twist – everyone knows that the second tier is inhabited entirely by witches, and therefore being on the second tier is sufficient to convict you. Unpopular ideas are gradually forced out of the first tier by media smear campaigns, and from then on everyone believes the effort was justified, because it's one of those second-tier ideas that you only find in the same sites as the racists and trolls and child pornographers. You're not a *second tier* kind of person, are you? No, we didn't think so.

I have no particular solution to this. Certainly the well-intentioned solutions other people are working on, like a decentralized crypto-Reddit that can't be moderated even in

principle, are unlikely to help (hint: what is the most striking difference between Bitcoin marketplaces and normal marketplaces?) My primary hope is that it's just not a real problem. Certainly there has been very little in the way of speech restriction so far, and what little there has been has been against things which, on the object level, I'm happy to see gone. It's entirely possible that we'll escape with only a few things banned that probably deserve it. I certainly hope this is the case.

I'm just annoyed that we've gotten ourselves in a corner where we have to depend on hope.

Book Review: Singer on Marx

I'm not embarassed for choosing Singer's Marx: A Very Short Introduction as a jumping-off point for learning more leftist philosophy. I weighed the costs and benefits of reading primary sources versus summaries and commentaries, and decided in favor of the latter.

The clincher was that the rare times I felt like I really understand certain thinkers and philosophies on a deep level, it's rarely been the primary sources that did it for me, even when I'd read them. It's only after hearing a bunch of different people attack the same idea from different angles that I've gotten the gist of it. The primary sources – especially when they're translated, especially when they're from the olden days before people discovered how to be interesting – just turn me off. Singer is a known person who can think and write clearly, and his book was just about the shortest I could find, so I jumped on it, hoping I would find a more sympathetic portrayal of someone whom my society has been trying to cast as a demon or monster.

And I don't know if this is an artifact of Singer or a genuine insight into Marx, but as far as I can tell he's even worse than I thought.

I.

What really clinched this for me was the discussion of Marx's (lack of) description of how to run a communist state. I'd always heard that Marx was long on condemnations of capitalism and short on blueprints for communism, and the couple of Marx's works I read in college confirmed he really

didn't talk about that very much. It seemed like a pretty big gap.

But I'd always dismissed this as an excusable error. When I was really young – maybe six or seven – I fancied myself a great inventor. The way I would invent something – let's say a spaceship – was to draw a picture of a spaceship. I would label it with notes like "engine goes here" and "power source here" and then rest on my laurels, satisfied that I had invented interstellar travel at age seven. It always confused me that adults, who presumably should be pretty smart, had failed to do this. Occasionally I would bring this up to someone like my parents, and they would ask a question like "Okay, but how does the power source work?" and I would answer "Through quantum!" and then get very annoyed that people *didn't even know about quantum*.

(I was seven years old. What's *your* excuse, New Age community?)

I figured that Marx had just fallen into a similar trap. He'd probably made a few vague plans, like "Oh, decisions will be made by a committee of workers," and "Property will be held in common and consensus democracy will choose who gets what," and felt like the rest was just details. That's the sort of error I could at least sympathize with, despite its horrendous consequences.

But in fact Marx was philosophically opposed, as a matter of principle, to any planning about the structure of communist governments or economies. He would come out and say "It is irresponsible to talk about how communist governments and economies will work." He believed it was a scientific law, analogous to the laws of physics, that once capitalism was removed, a perfect communist government would form of its

own accord. There might be some very light planning, a couple of discussions, but these would just be epiphenomena of the governing historical laws working themselves out. Just as, a dam having been removed, a river will eventually reach the sea somehow, so capitalism having been removed society will eventually reach a perfect state of freedom and cooperation.

Singer blames Hegel. Hegel viewed all human history as the World-Spirit trying to recognize and incarnate itself. As it overcomes its various confusions and false dichotomies, it advances into forms that more completely incarnate the World-Spirit and then moves onto the next problem. Finally, it ends with the World-Spirit completely incarnated – possibly in the form of early 19th century Prussia – and everything is great forever.

Marx famously exports Hegel's mysticism into a materialistic version where the World-Spirit operates upon class relations rather than the interconnectedness of all things, and where you don't come out and *call* it the World-Spirit – but he basically keeps the system intact. So once the World-Spirit resolves the dichotomy between Capitalist and Proletariat, then it can more completely incarnate itself and move on to the next problem. Except that this is the final problem (the proof of this is trivial and is left as exercise for the reader) so the World-Spirit becomes fully incarnate and everything is great forever. And you want to *plan* for how that should happen? Are you saying you know better than the World-Spirit, Comrade?

I am starting to think I was previously a little too charitable toward Marx. My objections were of the sort "You didn't really consider the idea of welfare capitalism with a social safety net" or "communist society is very difficult to implement in principle," whereas they should have looked

more like “You are basically just telling us to destroy all of the institutions that sustain human civilization and trust that what is *baaaasically* a giant planet-sized ghost will make sure everything works out.”

II.

Conservatives always complain that liberals “deny human nature”, and I had always thought that complaint was unfair. Like sure, liberals say that you can make people less racist, and one could counterargue that a tendency toward racism is inborn, but it sure seems like you can make that tendency more or less strongly expressed and that this is important. This is part of the view I argue in Nature Is Not A Slate, It’s A Series Of Levers.

But here I have to give conservatives their due. As far as I can tell, Marx literally, so strongly as to be unstrawmannable, believed there was no such thing as human nature and everything was completely malleable.

> Feuerbach resolves the essence of religion into the essence of man. But the essence of man is no abstraction inherent in each single individual. In reality, it is the ensemble of the social relations.

And:

> It is evidence that economics establishes an alienated form of social intercourse as the essential, original, and natural form

Which Singer glosses with:

> This is the gist of Marx’s objection to classical economics. Marx does not challenge the classical

> economists within the presuppositions of their science. Instead, he takes a viewpoint outside those presuppositions and argues that private property, competition, greed, and so on are to be found only in a particular condition of human existence, a condition of alienation.

I understand this is still a matter of some debate in the Marxist community. But it seems to me that if Singer is right, if this is genuinely Marx's view, it seems likely to be part of what contributed to his inexcusable error above.

You or I, upon hearing that the plan is to get rid of all government and just have people share all property in common, might ask questions like "But what if someone wants more than their share?" Marx had no interest in that question, because he believed that there was no such thing as human nature, and things like "People sometimes want more than their shares of things" are contingent upon material relations and modes of production, most notably capitalism. If you get rid of capitalism, human beings change completely, such that "wanting more than your share" is no more likely than growing a third arm.

A lot of the liberals I know try to distance themselves from people like Stalin by saying that Marx had a pure original doctrine that they corrupted. But I am finding myself much more sympathetic to the dictators and secret police. They may not have been very nice people, but they were, in a sense, operating in Near Mode. They couldn't just tell themselves "After the Revolution, no one is going to demand more than their share," because their philosophies were shaped by the experience of having their subordinates come up to them and say "Boss, that Revolution went great, but now someone's

demanding more than their share, what should we do?" Their systems seem to be part of the unavoidable collision of Marxist doctrine with reality. It's possible that there are other, better ways to deal with that collision, but "returning to the purity of Marx" doesn't seem like a workable option.

III.

There was one part that made me more sympathetic to Marx. Singer writes:

> Marx saw that the liberal definition of freedom is open to a fundamental objection. Suppose I live in the suburbs and work in the city. I could drive my car to work, or take the bus. I prefer not to wait around for the bus, and so I take my car. Fifty thousand other people living in my suburb face the same choice and make the same decision. The road to town is choked with cars. It takes each of us an hour to travel ten miles. In this situation, according to the liberal conception of freedom, we have all chosen freely. Yet the outcome is something none of us want. If we all went by bus, the roads would be empty and we could cover the distance in twenty minutes. Even with the inconvenience of waiting at the bus stop, we would all prefer that. We are, of course, free to alter our choice of transportation, but what can we do? While so many cars slow the bus down, why should any individual choose differently? The liberal conception of freedom has led to a paradox: we have each chosen in our own interests, but the result is in no one's interest. Individual rationality, collective irrationality…
>
> Marx saw that capitalism involves this kind of collective irrationality. In precapitalist systems it was obvious that most people did not control their own destiny – under

> feudalism, for instance, serfs had to work for their lords. Capitalism seems different because people are in theory free to work for themselves or for others as they choose. Yet most workers have as little control over their lives as feudal serfs. This is not because they have chosen badly, nor is it because of the physical limits of our resources and technology. It is because the cumulative effect of countless individual choices is a society that no one – not even the capitalists – has chosen. Where those who hold the liberal conception of freedom would say we are free because we are not subject to deliberate interference by other humans, Marx says we are not free because we do not control our own society.

This is good. In fact, this is the insight that I spent about fifteen years of my life looking for, ever since I first discovered libertarianism and felt like there was definitely an important problem with it, but couldn't quite verbalize what it was. It's something I finally figured out only within the last year or so and didn't fully write up until Meditations on Moloch. And Marx seems to have sort of had it. I read the relevant section of Marx when I was younger, where he was talking about how capitalists would compete each other into the ground whether they wanted to or not, and I remember dismissing it with a "capitalists have not competed each other into the ground, for this this and this reason", dismissing the incorrect object-level argument without realizing the important meta-level insight beneath it (something I have since learned to stop doing). If Marx really had that meta-level insight – really had it, and not just stumbled across a couple of useful examples of it without realizing the pattern – then that would make his fame justly deserved.

But two things here discourage me. First, Marx seems so confused about everything that it's hard to parse him as really understanding this, as opposed to simply noticing one example of it that serves as a useful argument against capitalism. I notice Singer had to come up with his own clever example of this instead of quoting anything from any of Marx's works. Second, the insight does not seem original to Marx. Tragedy of the commons was understood as early as 1833 and Malthus was talking about similar problems related to population explosions before Marx was even born. John Stuart Mill, writing twenty years before *Das Kapital*, had already explained the basic principle quite well:

> To a fourth case of exception I must request particular attention, it being one to which as it appears to me, the attention of political economists has not yet been sufficiently drawn. There are matters in which the interference of law is required, not to overrule the judgment of individuals respecting their own interest, but to give effect to that judgment: they being unable to give effect to it except by concert, which concert again cannot be effectual unless it receives validity and sanction from the law. For illustration, and without prejudging the particular point, I may advert to the question of diminishing the hours of labour. Let us suppose, what is at least supposable, whether it be the fact or not—that a general reduction of the hours of factory labour, say from ten to nine,*119 would be for the advantage of the workpeople: that they would receive as high wages, or nearly as high, for nine hours' labour as they receive for ten. If this would be the result, and if the operatives generally are convinced that it would, the limitation, some may say, will be adopted spontaneously. I answer,

> that it will not be adopted unless the body of operatives bind themselves to one another to abide by it. A workman who refused to work more than nine hours while there were others who worked ten, would either not be employed at all, or if employed, must submit to lose one-tenth of his wages. However convinced, therefore, he may be that it is the interest of the class to work short time, it is contrary to his own interest to set the example, unless he is well assured that all or most others will follow it. But suppose a general agreement of the whole class: might not this be effectual without the sanction of law? Not unless enforced by opinion with a rigour practically equal to that of law. For however beneficial the observance of the regulation might be to the class collectively, the immediate interest of every individual would lie in violating it: and the more numerous those were who adhered to the rule, the more would individuals gain by departing from it.

So one might apply to Marx the old cliche: that he has much that is good and original, but what is good is not original and what is original is not good.

But it is interesting to analyze Marx as groping toward something game theoretic. This comes across to me in some of his discussions of labor. Marx thinks all value is labor. Yes, capital is nice, but in a sense it is only “crystallized labor” – the fact that a capitalist owns a factory only means that at some other point he got laborers to build a factory for him. So labor does everything, but it gets only a tiny share of the gains produced. This is because capitalists are oppressing the laborers. Once laborers realize what’s up, they can choose to labor in such a way as to give themselves the full gains of their labor.

I think here that he is thinking of coordination as something that happens instantly in the absence of any obstacle to coordination, and the obstacle to coordination is the capitalists and the "false consciousness" they produce. Remove the capitalists, and the workers – who represent the full productive power of humanity – can direct that productive power to however it is most useful. In my language, Marx simply *assumed* the invisible nation, thought that the result of perfect negotiation by ideal game theoretic agents with 100% cooperation under a veil of ignorance – would also be the result of real negotiation in the real world, as long as there were no capitalists involved. Maybe this idea – of gradually approaching the invisible nation – is what stood in for the World-Spirit in his dialecticalism. Maybe in 1870, this sort of thinking was excusable.

If capitalists are to be thought of as anything other than parasites, part of the explanation of their contribution has to involve coordination. If Marx didn't understand that coordination is just as hard to produce as linen or armaments or whatever, if he thought you could just *assume* it, then capitalists seem useless and getting rid of all previous forms of government so that insta-coordination can solve everything seems like a pretty swell idea.

If you admit that, capitalists having disappeared, there's still going to be competition, positive and negative sum games, free rider problems, tragedies of the commons, and all the rest, then you've got to invent a system that solves all of those issues better than capitalism does. That seems to be the real challenge Marxist intellectuals should be setting themselves, and I hope to eventually discover some who have good answers to it. But at least from the little I learned from Singer,

I see no reason to believe Marx had the clarity of thought to even understand the question.

Does Class Warfare Have a Free Rider Problem?

Here are two comments I've gotten on this blog in the past few weeks:

> Progressivism is under massive selective pressure to actually cause problems because that leads to more power for progressivism.

> Sasha and Malia Obama will get affirmative action, even though their own father has publicly admitted its ridiculous. Therefore, black elites have a stake in keeping black masses as poor and miserable as possible, to continue justifying affirmative action.

These seem like they can be easily dismissed as conspiracy theories, but what is the exact structure of that dismissal?

Well, first, it requires that people have an almost comical level of evil. Think of the Secretary of Health and Human Services noticing that, if she enacted terrible policies that made everyone in the country sick, people would demand more resources for health care and her empire would grow. It's hard for me to imagine someone *that* Slytherin.

Second, it sounds like it requires literal conspiracy. In the second example, one of two things must happen. Either every black elite has to come up with the plan independently and work together in synchrony to carry it out – each taking it on faith that the other elites are doing their part. Or one person has to come up with the plan, convince everyone else that that's the plan, and send them their marching orders ("You! Do your part to help keep the masses poor by voting against this

much-needed education reform!"), all without the media catching wind of any of this.

Third, this makes the same mistake I accused Marx of in the last post. It assumes a free solution to all coordination problems.

Suppose we grant the conspiracy theorists their point that it is indeed in the interest of all black elites to keep the black masses poor so they can benefit from affirmative action. Suppose we even grant that they are evil enough to want to try this plan despite the suffering it will produce. And suppose they're all really good at communicating through heavily encrypted email, so we solve the conspiracy aspect. The plan *still* doesn't work.

Every elite benefits from the entire plan being pulled off. But now there's a free rider problem. Each elite would have to expend some individual effort to keep everybody else down. Maybe it's going out of their way to rally opposition to a useful reform. Maybe it's having to take an unpopular position and so looking like the bad guy. All I'm saying is that quashing the dreams of the next generation of minority children is harder than sitting on your *tuchus* playing video games. Their own contribution doesn't help the cause very much on net, so their incentive is to defect and hope everyone else does it.

Just as good people playing normal politics have a hard time rallying support for genuinely important causes like stopping global warming or enforcing Net Neutrality, so evil people playing Conspiracy Politics should have a hard time convincing their target demographic to get out of bed and join in their oppression.

But in fact they have it much harder. Good people playing normal politics can use a host of techniques – phone banks, door-to-door campaigns, benefit concerts, leaflets in the mail, celebrity endorsements – to rally people to action. Evil people playing Conspiracy Politics can't do any of that without greatly increasing their risk of getting caught.

And when good people do rally the masses to their cause, it seems to be through an appeal to morality. Like "Yes, I know it would be much easier for you to sit back and let other people solve global warming, but you have an *ethical responsibility* to participate in this, and won't you feel good about yourself knowing you've made a difference."

Obviously if your campaign is "Cause as many problems as possible to increase the size of government" this is harder to pull off.

This seems to me to be a little-acknowledged third reason to dismiss conspiracy theories of this sort. But you don't care. You've already wandered off, wondering why I'm wasting my time debunking things nobody (except apparently the rare SSC commenter) believes anyway.

But what if we apply this to more common claims? What about class warfare?

It is widely believed that the rich have captured government for their own ends. For example, rich people use their money and power to decrease tax rates on the wealthy and sabotage legislation meant to protect the working man.

But this ought to fall victim to the same coordination problems. After all, suppose you are a rich person who makes $1 million per year. You would like the government to cut federal taxes on the wealthy from 40% down to 30%, which

would save you $100,000 per year. One might think you would be willing to spend up to $100,000 to effect this goal.

But in fact it requires the concerted effort of all the rich people across the country to make this happen. A single $100,000 donation isn't going to change federal level policy in such a spectacular way. Realistically your effort will be a drop in a bucket that your entire class needs to contribute to.

Once again we encounter free rider problems. Suppose a representative of the Rich People's Union asks for a $10,000 donation to fight for lower taxes. There are hundreds of thousands of rich people, so you're pretty sure your one donation isn't going to push anything over the edge one way or the other. Supposing the tax cut goes through, you will get the same benefit whether you donated or not; supposing it doesn't, you won't gain anything either way. It's easy to see that in either case the rational self-interested thing to do is to refuse to donate.

There are a couple of rare exceptions to this. If you are Bill Gates and make a billion dollars a year, so that you would gain $100 million from the tax cut, it might be worth bribing the necessary legislators all on your own, on the grounds that if something needs to be done right you had better do it yourself. Likewise, if you're Exxon Mobil or the Koch brothers, then you might be a big enough chunk of the target population for certain specific environmental regulations that it's worth using your own money to fight it whether or not others join in.

But a general focus on the interests of the rich? Not likely.

Yet the rich do seem to get their way a disproportionate amount of the time, and this seems to require an explanation.

I am reminded of the research I looked at in Plutocracy Isn't About Money. People seem to donate surprisingly little to

political candidates, and donations don't seem to help. This seems consistent with the idea that rich people don't directly coordinate to bribe politicians in their favor. I suggested a couple of different hypotheses, like that maybe the rich win because of "soft power" – ie the media and universities and politicians are mostly rich or are run by rich people who just sort of naturally let their opinions percolate through without much deliberate effort.

An alternative explanation preserves our intuitive belief that the rich sure do seem to influence politics a lot. Maybe rich people, like poor people, participate in politics because of sincere belief in their moral values, and their values are by what seems a weird coincidence the ones that help make them richer.

Like, Mitt Romney's zillion-dollar-a-plate fundraisers seem to always be pretty full. It can't literally be in a rich person's self-interest to buy a plate there. But a lot of rich people could have conservative-libertarian-pro-business ideas that encourage them to quasi-altruistically support Mitt Romney in order to push their values.

But this is really weird and interesting – much more interesting than it looks. It suggests that, in the presence of a useful selfish goal to coordinate around, a value system will "spring up" that convinces people to support it for altruistic reasons.

I'm not just talking about normal altruism here. A rich person motivated by normal altruism per se might be against tax cuts for the rich, in order to better preserve social services for the less fortunate. And I'm not just talking about normal selfishness either. A rich person motivated by selfishness would hang out in his mansion all day instead of wasting

money on fundraisers. I'm talking about a moral system which is genuinely self-sacrificing on the individual level, but which when universalized has the effect of helping the rich person get richer.

It's worth thinking about this in contractarian terms. A rich person, minus the veil of ignorance, wouldn't support everyone pitching in to help the poor, because he knows he's not poor and so gains nothing. A rich person, minus the veil of ignorance, *would* support a binding pact among all rich people to pitch in to support tax cuts on the rich, because she knows she would gain more than she loses from such an agreement.

But as far as I can tell, this calculation is never made on a conscious level. What happens on a conscious level is the rich person finds themselves supporting some moral philosophy – libertarianism, Objectivism, prosperity gospel, whatever – which says it is morally wrong to raise taxes on the rich, so much so that one should altruistically make personal sacrifices in order to stop them from being raised. And then these moral philosophies spread, and without any conscious awareness, the rich people find themselves coordinating very nicely to protect their class interests.

I hope you agree that if this is true, it is *spooky*. I admit on this blog I sometimes mock human nature and human cognition a little too much, but this particular cognitive process is *really impressive*. I hope whatever angel designed it got a promotion.

So although I haven't really thought this through too much, I would suggest a dichotomy. Either there's some sort of spooky system that generates heartfelt moral philosophies on demand to solve coordination problems, or the rich aren't actually coordinating and just consistently keep getting lucky.

I don’t like this because it raises more questions than it answers. Why don’t the poor coordinate this well? Too many of them? And if this is true, how sure should we be of our previous belief that the Secretary of Health and Human Services isn’t coordinating with all the other progressive bureaucrats to deliberately cause social problems?

Book Review: Red Plenty

I.

I decided to read *Red Plenty* because my biggest gripe after reading Singer's book on Marx was that Marx refused to plan how communism would actually work, instead preferring to leave the entire matter for the World-Spirit to sort out. But almost everything that interests me about Communism falls under the category of "how communism would actually work". Red Plenty, a semi-fictionalized account of the history of socialist economic planning, seemed like a natural follow-up.

But I'd had it on my List Of Things To Read for even longer than that, ever after stumbling across a quote from it on some blog or other:

> Marx had drawn a nightmare picture of what happened to human life under capitalism, when everything was produced only in order to be exchanged; when true qualities and uses dropped away, and the human power of making and doing itself became only an object to be traded.
>
> Then the makers and the things made turned alike into commodities, and the motion of society turned into a kind of zombie dance, a grim cavorting whirl in which objects and people blurred together till the objects were half alive and the people were half dead. Stock-market prices acted back upon the world as if they were independent powers, requiring factories to be opened or closed, real human beings to work or rest, hurry or dawdle; and they, having given the transfusion that made the stock prices come

> alive, felt their flesh go cold and impersonal on them, mere mechanisms for chunking out the man-hours. Living money and dying humans, metal as tender as skin and skin as hard as metal, taking hands, and dancing round, and round, and round, with no way ever of stopping; the quickened and the deadened, whirling on.
>
> And what would be the alternative? The consciously arranged alternative? A dance of another nature. A dance to the music of use, where every step fulfilled some real need, did some tangible good, and no matter how fast the dancers spun, they moved easily, because they moved to a human measure, intelligible to all, chosen by all.

Needless to say, this is Relevant To My Interests, which include among them poetic allegories for coordination problems. And I was not disappointed.

II.

The book begins:

> Strange as it may seem, the gray, oppressive USSR was founded on a fairy tale. It was built on the twentieth-century magic called "the planned economy," which was going to gush forth an abundance of good things that the lands of capitalism could never match. And just for a little while, in the heady years of the late 1950s, the magic seemed to be working. Red Plenty is about that moment in history, and how it came, and how it went away; about the brief era when, under the rash leadership of Khrushchev, the Soviet Union looked forward to a future of rich communists and envious capitalists, when Moscow would out-glitter Manhattan and every Lada would be better engineered than a Porsche. It's about the

> scientists who did their genuinely brilliant best to make the dream come true, to give the tyranny its happy ending.

And this was the first interesting thing I learned.

There's a very settled modern explanation of the conflict between capitalism and communism. Capitalism is good at growing the economy and making countries rich. Communism is good at caring for the poor and promoting equality. So your choice between capitalism and communism is a trade-off between those two things.

But for at least the first fifty years of the Cold War, the Soviets would not have come *close* to granting you that these are the premises on which the battle must be fought. They were officially quite certain that any day now Communism was going to prove itself *better* at economic growth, better at making people rich quickly, than capitalism. Even unofficially, most of their leaders and economists were pretty certain of it. And for a little while, even their capitalist enemies secretly worried they were right.

The arguments are easy to understand. Under capitalism, plutocrats use the profits of industry to buy giant yachts for themselves. Under communism, the profits can be reinvested back into the industry to build more factories or to make production more efficient, increasing growth rate.

Under capitalism, everyone is competing with each other, and much of your budget is spent on zero-sum games like advertising and marketing and sales to give you a leg up over your competition. Under communism, there is no need to play these zero-sum games and that part of the budget can be reinvested to grow the industry more quickly.

Under capitalism, everyone is working against everyone else. If Ford discovers a clever new car-manufacturing technique, their first impulse is to patent it so GM can't use it, and GM's first impulse is to hire thousands of lawyers to try to thwart that attempt. Under communism, everyone is working together, so if one car-manufacturing collective discovers a new technique they send their blueprints to all the other car-manufacturing collectives in order to help them out. So in capitalism, each companies will possess a few individual advances, but under communism every collective will have every advance, and so be more productive.

These arguments make a lot of sense to me, and they *definitely* made sense to the Communists of the first half of the 20th century. As a result, they were confident of overtaking capitalism. They realized that they'd started with a handicap – czarist Russia had been dirt poor and almost without an industrial base – and that they'd faced a further handicap in having the Nazis burn half their country during World War II – but they figured as soon as they overcame these handicaps their natural advantages would let them leap ahead of the West in only a couple of decades. The great Russian advances of the 50s – Sputnik, Gagarin, etc – were seen as evidence that this was already starting to come true in certain fields.

And then it all went wrong.

III.

Grant that communism really does have the above advantages over capitalism. What advantage does capitalism have?

The classic answer is that during communism no one wants to work hard. They do as little as they can get away with, then slack off because they don't reap the rewards of their own labor.

Red Plenty doesn't really have theses. In fact, it's not really a non-fiction work at all. It's a dramatized series of episodes in the lives of Russian workers, politicians, and academics, intended to come together to paint a picture of how the Soviet economy worked.

But if I can impose a thesis upon the text, I don't think it agreed with this. In certain cases, Russians were *very* well-incentivized by things like "We will kill you unless you meet the production target". Later, when the state became less murder-happy, the threat of death faded to threats of demotions, ruined careers, and transfer to backwater provinces. And there were equal incentives, in the form of promotion or transfer to a desirable location such as Moscow, for overperformance. There were even monetary bonuses, although money bought a lot less than it did in capitalist countries and was universally considered inferior to status in terms of purchasing power. Yes, there were Goodhart's Law type issues going on – if you're being judged per product, better produce ten million defective products than 9,999,999 excellent products – but that wasn't the crux of the problem.

Red Plenty presented the problem with the Soviet economy primarily as one of allocation. You could have a perfectly good factory that could be producing lots of useful things if only you had one extra eensy-weensy part, but unless the higher-ups had allocated you that part, you were out of luck. If that part happened to break, getting a new one would depend on how much clout you (and your superiors) pulled versus how much clout other people who wanted parts (and their superiors) held.

The book illustrated this reality with a series of stories (I'm not sure how many of these were true, versus useful dramatizations). In one, a pig farmer in Siberia needed wood

in order to build sties for his pigs so they wouldn't freeze – if they froze, he would fail to meet his production target and his career would be ruined. The government, which mostly dealt with pig farming in more temperate areas, hadn't accounted for this and so hadn't allocated him any wood, and he didn't have enough clout with officials to request some. A factory nearby had extra wood they weren't using and were going to burn because it was too much trouble to figure out how to get it back to the government for re-allocation. The farmer bought the wood from the factory in an under-the-table deal. He was caught, which usually wouldn't have been a problem because *everybody* did this sort of thing and it was kind of the "smoking marijuana while white" of Soviet offenses. But at that particular moment the Party higher-ups in the area wanted to make an example of someone in order to look like they were on top of their game to *their* higher-ups. The pig farmer was sentenced to years of hard labor.

A tire factory had been assigned a tire-making machine that could make 100,000 tires a year, but the government had gotten confused and assigned them a production quota of 150,000 tires a year. The factory leaders were stuck, because if they tried to correct the government they would look like they were challenging their superiors and get in trouble, but if they failed to meet the impossible quota, they would all get demoted and their careers would come to an end. They learned that the tire-making-machine-making company had recently invented a new model that really *could* make 150,000 tires a year. In the spirit of Chen Sheng, they decided that since the penalty for missing their quota was something terrible and the penalty for sabotage was also something terrible, they might as well take their chances and destroy their own machinery in the hopes the government sent them the new improved machine as

a replacement. To their delight, the government believed their story about an "accident" and allotted them a new tire-making machine. *However*, the tire-making-machine-making company had decided to cancel production of their new model. You see, the new model, although more powerful, weighed less than the old machine, and the government was measuring their production *by kilogram of machine*. So it was easier for them to just continue making the old less powerful machine. The tire factory was allocated another machine that could only make 100,000 tires a year and was back in the same quandary they'd started with.

It's easy to see how all of these problems could have been solved (or would never have come up) in a capitalist economy, with its use of prices set by supply and demand as an allocation mechanism. And it's easy to see how thoroughly the Soviet economy was sabotaging itself by avoiding such prices.

IV.

The "hero" of *Red Plenty* – although most of the vignettes didn't involve him directly – was Leonid Kantorovich, a Soviet mathematician who thought he could solve the problem. He invented the technique of linear programming, a method of solving optimization problems perfectly suited to allocating resources throughout an economy. He immediately realized its potential and wrote a nice letter to Stalin politely suggesting his current method of doing economics was wrong and he could do better – this during a time when everyone else in Russia was desperately trying to avoid having Stalin notice them because he tended to kill anyone he noticed. Luckily the letter was intercepted by a kindly mid-level official, who kept it away from Stalin and warehoused Kantorovich in a university somewhere.

During the “Khruschev thaw”, Kantorovich started getting some more politically adept followers, the higher-ups started taking note, and there was a real movement to get his ideas implemented. A few industries were run on Kantorovichian principles as a test case and seemed to do pretty well. There was an inevitable backlash. Opponents accused the linear programmers of being capitalists-in-disguise, which wasn’t helped by their use of something called “shadow prices”. But the combination of their own political adeptness and some high-level support from Khruschev – who alone of all the Soviet leaders seemed to really believe in his own cause and be a pretty okay guy – put them within arm’s reach of getting their plans implemented.

But when elements of linear programming were adopted, they were adopted piecemeal and toothless. The book places the blame on Alexei Kosygen, who implemented a bunch of economic reforms that failed, in a chapter that makes it clear exactly how constrained the Soviet leadership really was. You hear about Stalin, you imagine these guys having total power, but in reality they walked a narrow line, and all these “shadow prices” required more political capital than they were willing to mobilize, even when they thought Kantorovich might have a point.

V.

In the end, I was left with two contradictory impressions from the book.

First, amazement that the Soviet economy got as far as it did, given how incredibly screwed up it was. You hear about how many stupid things were going on at every level, and you think: *This was the country that built Sputnik and Mir? This was the country that almost buried us beneath the tide of*

history? It is a credit to the Russian people that they were able to build so much as a screwdriver in such conditions, let alone a space station.

But second, a sense of what could have been. What if Stalin *hadn't* murdered most of the competent people? What if entire fields of science *hadn't* been banned for silly reasons? What if Kantorovich *had* been able to make the Soviet leadership base its economic planning around linear programming? How might history have turned out differently?

One of the book's most frequently-hammered-in points was that there was was a brief moment, back during the 1950s, when everything seemed to be going right for Russia. Its year-on-year GDP growth (as estimated by impartial outside observers) was somewhere between 7 to 10%. Starvation was going down. Luxuries were going up. Kantorovich was fixing entire industries with his linear programming methods. Then Khruschev made a serious of crazy loose cannon decisions, he was ousted by Brezhnev, Kantorovich was pushed aside and ignored, the "Khruschev thaw" was reversed and tightened up again, and everything stagnated for the next twenty years.

If Khruschev had stuck around, if Kantorovich had succeeded, might the common knowledge that Communism is terrible at producing material prosperity look a little different?

The book very briefly mentioned a competing theory of resource allocation promoted by Victor Glushkov, a cyberneticist in Ukraine. He thought he could use computers – then a very new technology – to calculate optimal allocation for everyone. He failed to navigate the political seas as adroitly as Kantorovich's faction, and the killing blow was a paper that pointed out that for him to do everything *really*

correctly would take a hundred million years of computing time.

That was in 1960. If computing power doubles every two years, we've undergone about 25 doubling times since then, suggesting that we ought to be able to perform Glushkov's calculations in three years – or three days, if we give him a lab of three hundred sixty five computers to work with. There could have been this entire field of centralized economic planning. Maybe it would have continued to underperform prices. Or maybe after decades of trial and error across the entire Soviet Union, it could have caught up. We'll never know. Glushkov and Kantorovich were marginalized and left to play around with toy problems until their deaths in the 80s, and as far as I know their ideas were never developed further in the context of a national planned economy.

VI.

One of the ways people like insulting smart people, or rational people, or scientists, is by telling them they're the type of people who are attracted to Communism. "Oh, you think you can control and understand everything, just like the Communists did."

And I had always thought this was a pretty awful insult. The people I know who most identify as rationalists, or scientifically/technically minded, are also most likely to be libertarian. So there, case dismissed, everybody go home.

This book was the first time that I, as a person who considers himself rationally/technically minded, realized that I was super attracted to Communism.

Here were people who had a clear view of the problems of human civilization – all the greed, all the waste, all the zero-sum games. Who had the entire population united around a

vision of a better future, whose backers could direct the entire state to better serve the goal. All they needed was to solve the engineering challenges, to solve the equations, and there they were, at the golden future. And they were smart enough to be worthy of the problem – Glushkov invented cybernetics, Kantorovich won a Nobel Prize in Economics.

And in the end, they never got the chance. There's an interpretation of Communism as a refutation of social science, here were these people who probably knew some social science, but did it help them run a state, no it didn't. But from the little I learned about Soviet history from this book, this seems diametrically wrong. The Soviets had practically no social science. They hated social science. You would think they would at least have some good Marxists, but apparently Stalin killed all of them just in case they might come up with versions of Marxism he didn't like, and in terms of a vibrant scholarly field it never recovered. Economics was tainted with its association with capitalism from the very beginning, and when it happened at all it was done by non-professionals. Kantorovich was a mathematician by training; Glushkov a computer scientist.

Soviet Communism isn't what happens when you let nerds run a country, it's what happens when you kill all the nerds who are experts in country-running, bring in nerds from unrelated fields to replace them, then make nice noises at those nerds in principle while completely ignoring them in practice. Also, you ban all Jews from positions of importance, because fuck you.

> Baggy two-piece suits are not the obvious costume for philosopher kings: but that, in theory, was what the apparatchiks who rule the Soviet Union in the 1960s were

supposed to be. Lenin's state made the same bet that Plato had twenty-five centuries earlier, when he proposed that enlightened intelligence gives absolute powers would serve the public good better than the grubby politicking of republics.

On paper, the USSR was a republic, a grand multi-ethnic federation of republics indeed and its constitutions (there were several) guaranteed its citizens all manner of civil rights. But in truth the Soviet system was utterly unsympathetic to the idea of rights, if you meant by them any suggestion that the two hundred million men, women and children who inhabited the Soviet Union should be autonomously fixing on two hundred million separate directions in which to pursue happiness. This was a society with just one programme for happiness, which had been declared to be scientific and therefore was as factual as gravity.

But the Soviet experiment had run into exactly the difficulty that Plato's admirers encountered, back in the fifth century BC, when they attempted to mould philosophical monarchies for Syracuse and Macedonia. The recipe called for rule by heavily-armed virtue—or in the Leninist case, not exactly virtue, but a sort of intentionally post-ethical counterpart to it, self-righteously brutal. Wisdom was to be set where it could be ruthless. Once such a system existed, though, the qualities required to rise in it had much more to do with ruthlessness than wisdom. Lenin's core of Bolsheviks, and the socialists like Trotsky who joined them, were many of them highly educated people, literate in multiple European languages, learned in the scholastic traditions of Marxism; and they preserved these attributes even as

> they murdered and lied and tortured and terrorized. They were social scientists who thought principle required them to behave like gangsters. But their successors – the vydvizhentsy who refilled the CEntral Committee in the thirties – were not the most selfless people in Soviet society, or the most principled, or the most scrupulous. They were the most ambitious, the most domineering, the most manipulative, the most greedy, the most sycophantic: people whose adherence to Bolshevik ideas was inseparable from the power that came with them. Gradually their loyalty to the ideas became more and more instrumental, more and more a matter of what the ideas would let them grip in their two hands…
>
> Stalin had been a gangster who really believed he was a social scientist. Khruschev was a gangster who hoped he was a social scientist. But the moment was drawing irresistibly closer when the idealism would rot away by one more degree, and the Soviet Union would be governed by gangsters who were only pretending to be social scientists.

And in the end it all failed miserably:

> The Soviet economy did not move on from coal and steel and cement to plastics and microelectronics and software design, except in a very few military applications. It continued to compete with what capitalism had been doing in the 1930s, not with what it was doing now. It continued to suck resources and human labour in vast quantities into a heavy-industrial sector which had once been intended to exist as a springboard for something else, but which by now had become its own justification. Soviet industry in its last decades existed because it

> existed, an empire of inertia expanding ever more slowly, yet attaining the wretched distinction of absorbing more of the total effort of the economy that hosted it than heavy industry has ever done anywhere else in human history, before or since. Every year it produced goods that less and less corresponded to human needs, and whatever it once started producing, it tended to go on producing ad infinitum, since it possessed no effective stop signals except ruthless commands from above, and the people at the top no longer did ruthless, in the economic sphere. The control system for industry grew more and more erratic, the information flowing back to the planners grew more and more corrupt. And the activity of industry , all that human time and machine time it used up, added less and less value to the raw materials it sucked in. Maybe no value. Maybe less than none. One economist has argued that, by the end, it was actively destroying value; it had become a system for spoiling perfectly good materials by turning them into objects no one wanted.

I don't know if this paragraph was intentionally written to contrast with the paragraph at the top, the one about the zombie dance of capitalism. But it is certainly instructive to make such a contrast. The Soviets had originally been inspired by this fear of economics going out of control, abandoning the human beings whose lives it was supposed to improve. In capitalist countries, people existed for the sake of the economy, but under Soviet communism, the economy was going to exist only for the sake of the people.

(accidental Russian reversal: the best kind of Russian reversal!)

And instead, they ended up taking “people existing for the sake of the economy” to entirely new and tragic extremes, people being sent to the gulags or killed because they didn’t meet the targets for some product nobody wanted that was listed on a Five-Year Plan. Spoiling good raw materials for the sake of being able to tell Party bosses and the world “Look at us! We are doing Industry!” Moloch had done some weird judo move on the Soviets’ attempt to destroy him, and he had ended up stronger than ever.

The book’s greatest flaw is that it never did get into the details of the math – or even more than a few-sentence summary of the math – and so I was left confused as to whether anything else had been possible, whether Kantorovich and Glushkov really could have saved the vision of prosperity if they’d been allowed to do so. Nevertheless, the Soviets earned my sympathy and respect in a way Marx so far has not, merely by acknowledging that the problem existed and through the existence of a few good people who tried their best to solve it.