



Slate Star Codex Abridged

Welcome to SSC Abridged



This is a collection of some of the best essays written by psychiatrist and author Scott Alexander on his long-running blog, [Slate Star Codex](#). They're divided into several categories, such as "Science and Pseudoscience", "Liberalism and Its Enemies", "Medicine", etc.

Slate Star Codex is currently inactive (visit the blog to read about the reasons for this). As a result, finding Alexander's best writing is difficult, and reading his essays on *SSC* in its current form is inconvenient. Fortunately, all the content of *Slate Star Codex* is licensed under a [Creative Commons license](#) that allows us to re-release Scott's work in this format.

(The "abridged" in this book's title doesn't, by the way, mean that any of the posts have been shortened—only that the collection as a whole is a selected subset of Scott Alexander's writing. Each individual post comes to you in its full and glorious length; not a word has been omitted from any of these essays.)

Slate Star Codex Abridged is licensed under [the same license \(CC BY\)](#) as *Slate Star Codex*. (Our contribution to the content herein may be attributed to "Slate Star Codex Abridged".)

Note on dead links

In the years since these essays have been written, many of the links in the *Slate Star Codex* posts included here have, unfortunately, succumbed to link rot. These are gradually being fixed—replaced with archived versions where possible, etc.—but the process is slow. We apologize for the inconvenience in the meantime.

Liberalism And its Enemies

In Favor of Niceness, Community, and Civilization...12
Promising The Moon...24
Reactionary Philosophy In An Enormous Planet Sized Nutshell...28
Archipelago and Atomic Communitarianism...56
Book Review: On The Road...67
The Anti-Reactionary FAQ...73

Moloch

- [REPOST] The Non-Libertarian FAQ...158
- Growing Children For Bostrom's Disneyland...211
 - Book Review: Age of Em...216
 - Meditations On Moloch...230
 - Book Review: Red Plenty...258
 - The Goddess of Everything Else...265

Culture War

The Toxoplasma Of Rage...270
Beware The Man Of One Study...284
Infinite Debt...290
Bottomless Pits Of Suffering...294
Nydwracu's Fnords...296
The Virtue of Silence...300
Fundamental Value Differences Are Not That Fundamental...302
Beware Isolated Demands For Rigor...308
Society Is Fixed, Biology Is Mutable...313
The Parable Of The Talents...315
I Can Tolerate Anything Except The Outgroup...324
Can Things Be Both Popular And Silenced?...339
Social Censorship: The First Offender Model...349
Vote On Values, Outsource Beliefs...351
Weak Men Are Superweapons...354
How Did New Atheism Fail So Miserably?...360
The Influenza Of Evil...363
Guided By The Beauty Of Our Weapons...367
Ethics Offsets...376
Sort By Controversial...379

Science and Pseudoscience

- [REPOST] Epistemic Learned Helplessness...386
- If It's Worth Doing, It's Worth Doing With Made-Up Statistics...389
 - Lizardman's Constant Is 4%...392
 - The Control Group Is Out Of Control...395
 - Meat Your Doom...405
- SSC Journal Club: Expert Prediction Of Experiments...409
 - My IRB Nightmare...412
- Book Review: The Structure Of Scientific Revolutions...423

Psychiatry, Psychology, & Neurology

- The Phatic And The Anti-Inductive...433
- CBT In The Water Supply...438
- What Universal Human Experiences Are You Missing Without Realizing It?...444
 - What Developmental Milestones Are You Missing?...447
- An Iron Curtain Has Descended Upon Psychopharmacology...451
 - Why Were Early Psychedelicists So Weird?...454
 - Universal Love, Said The Cactus Person...457
- Did A Melatonin Patent Inspire Current Dose Confusion?...463
 - It's Bayes All The Way Up...465
- Book Review: Surfing Uncertainty...475
- Book Review: Behavior – The Control Of Perception...488
- Predictive Processing And Perceptual Control...494

Medicine

Who By Very Slow Decay...497

Book Review: House of God...505

Side Effects May Include Anything...514

Cost Disease In Medicine: The Practical Perspective...516

The Amish Health Care System...518

Sociology and Economics

The Ideology Is Not The Movement...525
Against Tulip Subsidies...536
Book Review: Hive Mind...541
SSC Gives A Graduation Speech...550
Basic Income, Not Basic Jobs: Against Hijacking Utopia...559
Bulls**t Jobs (Part 1 of ∞)...583
Considerations On Cost Disease...585
What Happened To '90s Environmentalism?...604
The Atomic Bomb Considered As Hungarian High School Science Fair Project...626
Hungarian Education II: Four Nobel Truths...634
Hungarian Education III: Mastering The Core Teachings Of The Budapestians...638
Book Review: Seeing Like A State...643
Book Review: Legal Systems Very Different From Ours...664
Book Review: The Secret Of Our Success...673
Book Review: Secular Cycles...687
Book Review: Ages Of Discord...700
1960: The Year The Singularity Was Cancelled...719
Ars Longa, Vita Brevis...731

“Powers, complete mental revision, ultraintelligence, posthumanity...”

Book Review and Highlights: Quantum Computing Since Democritus...739

Answer to Job...744

Wirehead Gods On Lotus Thrones...747

...And I Show You How Deep The Rabbit Hole Goes...750

SSC Journal Club: AI Timelines...764

Asches to Asches...770

SSC Journal Club: Dissolving The Fermi Paradox...776

Book Review: Reframing Superintelligence...780

It Was You Who Made My Blue Eyes Blue...784

Liberalism And its Enemies



“ *I disapprove of what you say, but I will defend to the death your right to say it.*

,

— Evelyn Beatrice Hall (often attributed to Voltaire)

**

[In Favor of Niceness, Community, and Civilization](#)

[Promising The Moon](#)

[Reactionary Philosophy In An Enormous Planet Sized Nutshell](#)

[Archipelago and Atomic Communitarianism](#)

[Book Review: On The Road](#)

[The Anti-Reactionary FAQ](#)

In Favor of Niceness, Community, and Civilization

Posted on February 23, 2014 by Scott Alexander



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. I've renamed him here to "Andrew Cord" to protect his identity.

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 “trigge~~r~~ry” 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.

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 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'm 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.

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 the guy in the fake rape statistics post 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 their 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 on 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 equanimity, 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 looked over the false rape statistics article and decided to lock 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 newspapers, 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 Alzheimer's.

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 on Facebook 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!”

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 knowledge 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)

Promising The Moon

Posted on July 21, 2014 by Scott Alexander



Note 11/19/19: In *Promising The Moon* (written 2014), I argued that technological progress had not slowed from 1970 to the present (compared to its pre-1970 speed). Since then, research by a bunch of people (summarized very well [here](#) by Tyler Cowen) finds strong evidence that it has. Despite its confidence and snark, my 2014 post was wrong.

I

The year 1969 comes up to you and asks what sort of marvels you've got all the way in 2014.

You explain that cameras, which 1969 knows as bulky boxes full of film that takes several days to get developed in dark rooms, are now instant affairs of point-click-send-to-friend that are also much higher quality. Also they can take video.

Music used to be big expensive records, and now you can fit 3,000 songs on an iPod and get them all for free if you know how to pirate or scrape the audio off of YouTube.

Television not only has gone HDTV and plasma-screen, but your choices have gone from “whatever’s on now” and “whatever is in theaters” all the way to “nearly every show or movie that has ever been filmed, whenever you want it”.

Computers have gone from structures filling entire rooms with a few kB memory and a punchcard-based interface, to small enough to carry in one hand with a few Tb memory and a touchscreen-based interface. And they now have peripherals like printers, mice, scanners, and flash drives.

Lasers have gone from only working in special cryogenic chambers to working at room temperature to fitting in your pocket to being ubiquitous in things as basic as supermarket checkout counters.

Telephones have gone from rotary-dial wire-connected phones that still sometimes connected to switchboards, to cell phones that fit in a pocket. But even better is bypassing them entirely and making video calls with anyone anywhere in the world for free.

Robots now vacuum houses, mow lawns, clean office buildings, perform surgery, participate in disaster relief efforts, and drive cars better than humans. Occasionally if you are a bad person a robot will swoop down out of the sky and kill you.

For better or worse, video games now exist.

Medicine has gained CAT scans, PET scans, MRIs, lithotripsy, liposuction, laser surgery, robot surgery, and telesurgery. Vaccines for pneumonia, meningitis, hepatitis, HPV, and chickenpox. Ceftriaxone, furosemide, clozapine, risperidone, fluoxetine, ondansetron, omeprazole, naloxone, suboxone, mefloquine, – and for that matter Viagra. Artificial hearts, artificial livers, artificial cochleae, and artificial legs so good that their users can compete in the Olympics. People with artificial eyes can only identify vague shapes at best, but they’re getting better every year.

World population has tripled, in large part due to new agricultural advantages. Catastrophic disasters have become much rarer, in large part due to architectural advances and satellites that can watch the weather from space.

We have a box which you can type something into and it will tell you everything anyone has ever written relevant to your query.

We have a place where you can log into from anywhere in the world and get access to approximately all human knowledge, from the scores of every game in the [1956 Roller Hockey World Cup](#) to [85 different side effects](#) of an obsolete antipsychotic medication. It is all searchable instantaneously. Its main problem is that people try to add so much information to it that its (volunteer) staff are constantly busy deleting information that might be extraneous.

We have the ability to translate nearly major human language to any other major human language instantaneously at no cost with relatively high accuracy.

We have navigation technology that over fifty years has gone from “map and compass” to “you can say the name of your destination and a small box will tell you step by step which way you should be going”.

We have the aforementioned camera, TV, music, videophone, video games, search engine, encyclopedia, universal translator, and navigation system all bundled together into a small black rectangle that fits in your pockets, responds to your spoken natural-language commands, and costs so little that Ethiopian subsistence farmers routinely use them to sell their cows.

But, you tell 1969, we have something more astonishing still. Something even more unimaginable.

“We have,” you say, “people who believe technology has stalled over the past forty-five years.”

1969’s head explodes.

II

It’s the anniversary of the moon landing, which means I have to deal with people passing around memes like this:



But I probably can't blame the date for the recent discussion here of [whether technological progress halted in 1972](#).

So I would like to take a moment to critique a certain strain of futurology.

There seems to be this thing where people imagine something that would look really cool, and predict that if we work hard on it for fifty years, we'll be able to pull it off. And then fifty years later, when barely any work has been done on it at all, they start looking for someone to blame.

Missions to Mars. Lunar colonies. Giant floating solar power satellites. Undersea domes. Ten mile high arcologies. Humanoid robots.

Whereas real technology doesn't advance by heading in the direction of something that looks cool, unless some government or tycoon is throwing lots of money in the direction of coolness. Real technology hill-climbs towards things that are useful and profitable.

Why haven't we colonized space yet? For the same reason we haven't colonized Antarctica. It's very cold and not a lot of fun and if you go outside you die.

In fact, Antarctica is preferable to space in pretty much every way. There is no reason to colonize space before you have finished colonizing Antarctica. And there is no reason to colonize Antarctica until you have finished colonizing Nebraska (population: 9 people per square km).

I will maintain that even if we had enough space flight technology that elementary school classes routinely took field trips to Mars, Mars would end up with two or three scientific bases, a resort where tourists could take their pictures on Olympus Mons, a compound of very dedicated libertarians, and nothing else. No domed cities. No colonies fighting for independence. Think that's implausible? School children take field trips to the Mojave Desert all the time, and it pretty much looks like that. Why should Mars prosper more than a much more habitable comparison area?

Likewise, the reason we don't build undersea domes isn't because we're not good enough. It's because humans breath better on land, and there's still a lot of land left to live in. On the rare occasion we want a resource located underwater, we build an oil rig on top of it and pump it from the surface, ie the part of the ocean where you don't get insta-crushed by ten atmospheres of pressure if something goes wrong.

And the reason there are no ten-mile-high arcologies is that we haven't already tiled all the desirable real estate with 9.9-mile-high arcologies and decided we still need more space.

Science fiction authors and would-be prophets stubbornly refuse to admit "would anybody reasonably pay money for this?" into their calculations. And so every ten years they end up predicting the "smart house". Where from your phone, you can control the lights in any room of the house! I imagine futurologists sitting in their kitchens, thinking "Oh no! I wish the lights were on in my bedroom, but all I have is my phone!" Maybe one day we will have houses that contain teleporters that can bring to any other building in the world without stepping outside. But if you're in the kitchen and you want the light on in the teleporter room, you'll still just walk to the teleporter room and flip the @#\$%ing switch.

I am not defending this as a normative view of how progress should work. There is a lot to be said for colonizing Mars as a survival strategy in case something unexpected happens to Earth. And there's also a lot to be said for Manhattan Project style efforts to discover a technology in a non-hill-climbing way, something where there's not a profitable transitional form at each step between where we are and what we want. But I would suggest we stick to those criticisms, and not to a criticism of advance per se.

(actually, we're not even all that bad at getting past the hill-climbing thing; government subsidies to solar seem to have been a very successful attempt to push solar out of an area where it wasn't profitable to improve into an area where it is)

But it's going to take some pretty creative accounting to make moon shots profitable. The main reason people funded the moon landing in 1969 (as opposed to the reason that people not involved in funding felt good about it) was to beat Russia and then get to rub it in their face forever. Nowadays that's no longer so fun (although rapidly becoming funner!). Therefore, we get the expected outcome of fewer moon shots until someone else thinks of a compelling incentive to go to the moon. So far there isn't one. There's no need to bring technological stagnation into the picture.

Reactionary Philosophy In An Enormous Planet Sized Nutshell

Posted on March 3, 2013 by Scott Alexander



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 present 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 constitutional 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’mom, 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 kind 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.

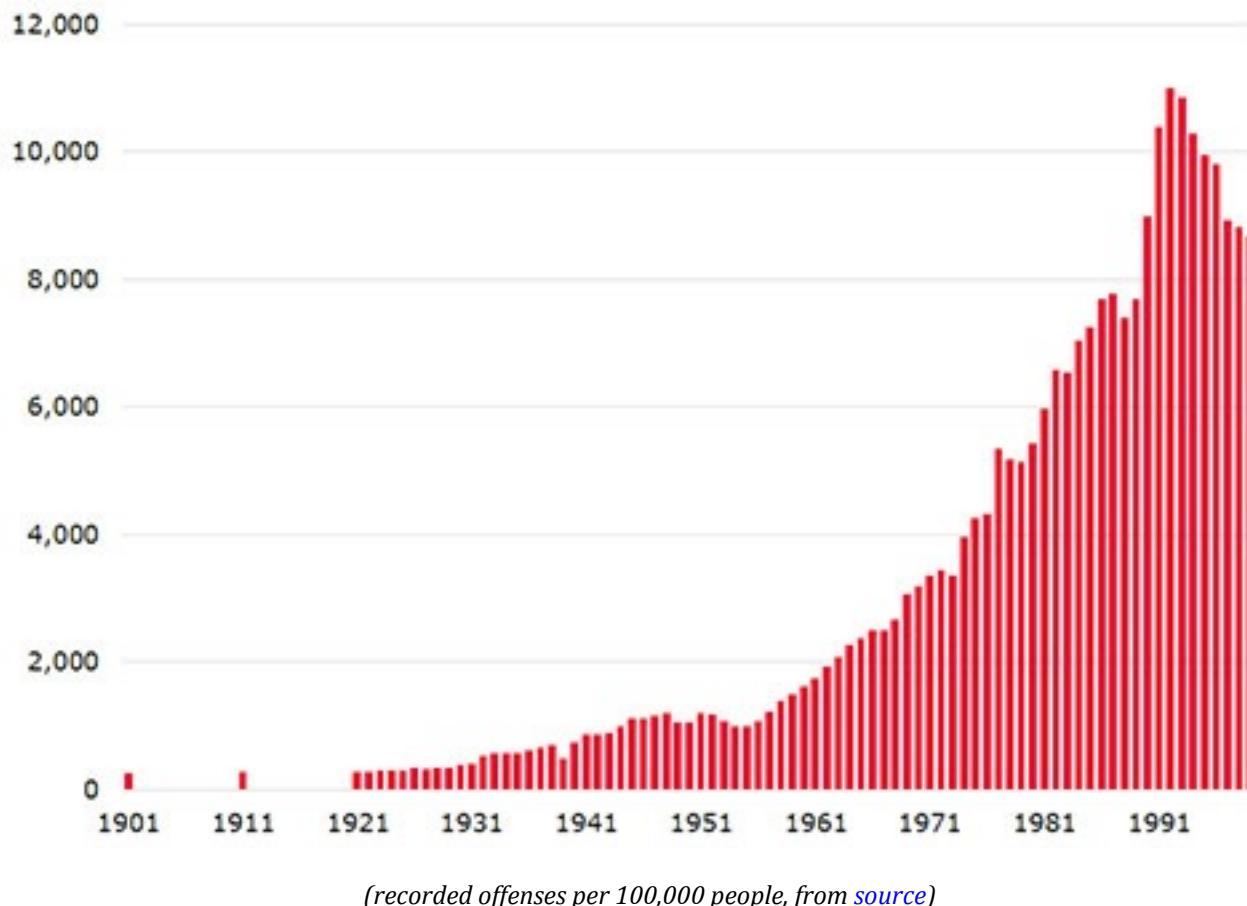
This technology dividend appears even in unexpected places. The world is more peaceful today, but how much of that is the existence of global trade networks that make war unprofitable, video reporting of every casualty that makes war unpopular, and nuclear and other weapons that make war unwinnable?

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:



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 humanness. 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.

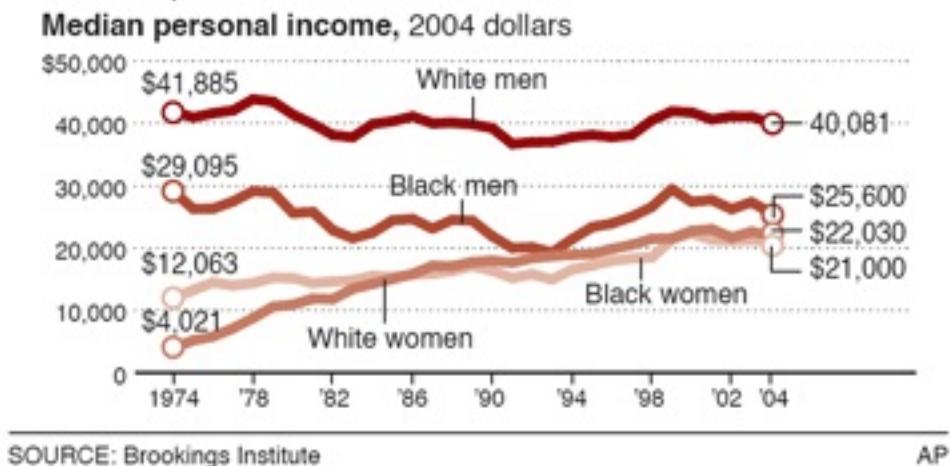
Like the externalists, they trace this forward to the present, saying that the values that served the Chinese so well in building Chinese civilization are the same ones that keep China strong today and the ones that make Chinese immigrants successful in countries like Malaysia and the USA. On the other 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?

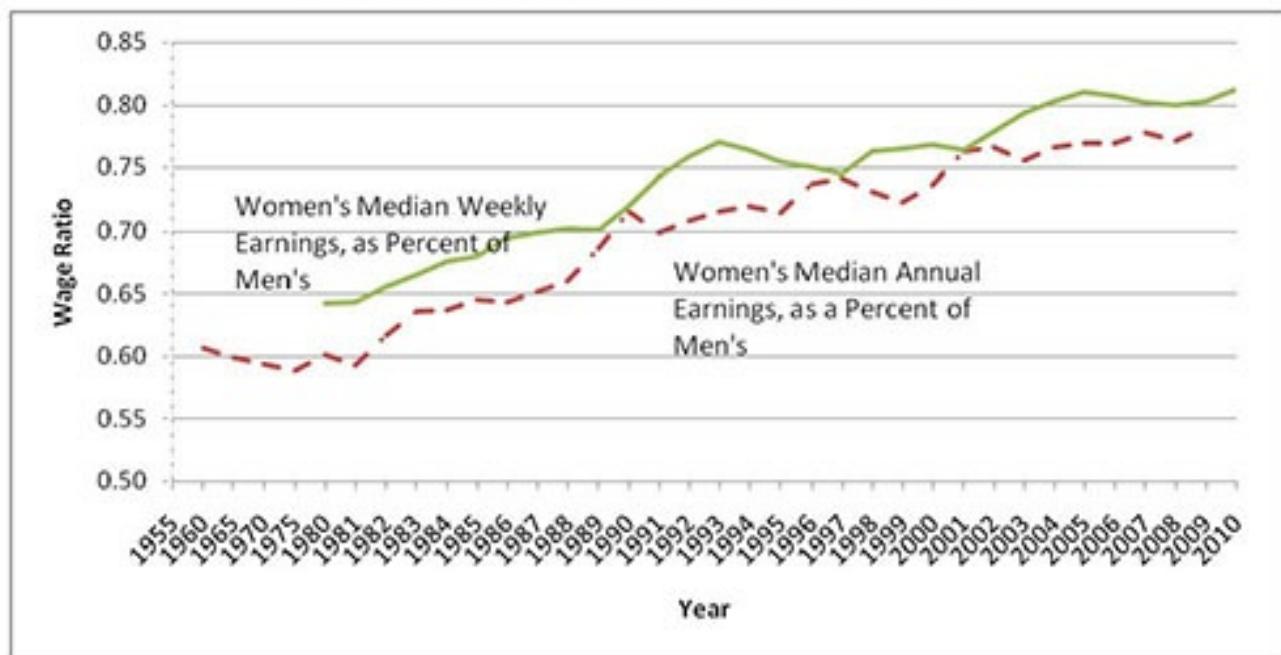
A decent amount of political wrangling over the years seems to involve a conflict between the conservatives – who are some vague mix of the culturalist and biologicalist position – and the liberals, who have embraced the externalist position with gusto.

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:

The Gender Earnings Ratio, 1955-2010, Full-Time Workers



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 memplexes they happened to end out with. Change the memplexes 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 remained 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 outweigh 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 probably 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 this 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 condition 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* than 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 annihilate 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 individuality – 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 decuple 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 thoroughly.

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 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 with 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.

Archipelago and Atomic Communitarianism

Posted on June 7, 2014 by Scott Alexander



I

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 maybe just liberalism boiled down to its pure essence.

But atomic individualism wasn't as great a solution as it sounded. Maybe one of the first cracks 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 of my coworkers brought in a box of free donuts and placed wide-open in the doctors' lounge; 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 even if they did I'd just encounter the wide-open box of free donuts *somewhere else*. I'm 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 atomic individualism as construed above is at best an imperfect heuristic.

And I would be remiss talking about the modern turn away from individualism without mentioning social justice. The same people who once deployed individualistic 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, pornographic media makes women hell 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 [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 if does.

II

A while ago, in Part V of [this essay](#), I praised liberalism as the only peaceful answer to Hobbes’ dilemma of the war of all against all.

Hobbes says that if everyone’s fighting then 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 crushes them like a bug, no prolonged fighting involved.

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 sort.

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 I, 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.

III

Imagine a new frontier suddenly opening. Maybe a wizard appears and gives us a map to a new archipelago that

geographers had missed for the past few centuries. He doesn't want to rule the archipelago himself, though he will reluctantly help kickstart the government. He just wants to give directions and a free galleon to anybody who wants one and can muster a group of likeminded friends large enough to start a self-sustaining colony.

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 the wizard decides to formalize and strengthen his system, not to mention work out some of the ethical dilemmas.

First he bans 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 united government (UniGov for short). 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.

Next he addresses externalities. For example, if some communities emit a lot of carbon, and that causes global warming which threatens to destroy other communities, UniGov puts a stop to that. If the offending communities refuse to stop emitting carbon, then there's that military again.

The third thing he 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 at it. 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 continent, and no one is allowed to force them to have more than that.

But the wizard and UniGov's most important task 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 the next island over. The *absolute minimum* the united 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 UniGov 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, UniGov burns their city and salts their field.

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 UniGov 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 UniGov representative in which they learn about the other communities in the Archipelago, receive a basic non-brainwashed view of the world, and are given directions to their nearest UniGov representative who they can give their opt-out request to.

The list of communities they are informed about always starts with the capital, ruled by UniGov itself and 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, UniGov will provide them with a galleon to go out and found their own community in hitherto uninhabited lands.

There's one more problem UniGov has to deal with: 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 the capital, which has an official commitment 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 UniGov 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 UniGov itself, just as a way of disincentivizing malicious movement.

Because there *are* UniGov taxes, and most people are happy to pay them. In my fantasy, UniGov 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 barbaric outside world, and it is precisely the institution that allows them to maintain their distinctiveness in the face of what would otherwise be irresistible 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.

IV

Robert Nozick once proposed a similar idea as a libertarian utopia, and it's easy to see why. UniGov 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 United Government of Archipelago is the perfect monarchist night watchman state, and any additions you make over that are chosen by your own free will.

But other people could view the same plan as 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 no one would 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 made 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"?

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.

V

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 of 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, redistributive 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.

Book Review: On The Road

Posted on December 2, 2014 by Scott Alexander



I

On the Road is a terrible book about terrible people. Jack Kerouac and his terrible friends drive across the US about seven zillion times for no particular reason, getting in car accidents and stealing stuff and screwing women whom they promise to marry and then don't.

But this is supposed to be okay, because they are *visionaries*. Their vision is to use the words "holy", "ecstatic", and "angelic" at least three times to describe every object between Toledo and Bakersfield. They don't pass a *barn*, they pass a holy vision of a barn, a barn such as there must have been when the world was young, a barn whose angelic red and beatific white send them into mad ecstasies. They don't almost hit a *cow*, they almost hit a holy primordial cow, the cow of all the earth, the cow whose dreamlike ecstatic mooing brings them to the brink of a rebirth such as no one has ever known.

Jack Kerouac and his terrible friends are brought to the brink of a lot of things, actually. Aside from stealing things and screwing women whom they promise to marry and then don't, being brought to the brink of things is one of their main pastimes. Enlightenment, revelation, truth, the real meaning of America, the ultimate, the sacred – if it has a brink, they will come to it. Crucially, they never cross that brink or gain any lasting knowledge or satisfaction from the experience. Theirs is a religion whose object of worship is the burst of intense emotion, the sudden drenching of their brain in happy chemicals that come and go without any lasting effect except pages full of the words "holy", "ecstatic", and "angelic".

The high priest of this religion is Kerouac's friend Dean Moriarty. Kerouac cannot frickin shut up about Dean Moriarty. Obviously he is "holy" and "ecstatic" and "angelic" and "mad" and "visionary", but for Dean, Kerouac pulls out all the stops. He is "a new kind of American saint", "a burning shuddering frightful Angel", with intelligence "formal and shining and complete".

Who is this superman, this hero?

His specialty was stealing cars, gunning for girls coming out of high school in the afternoon, driving them out to the mountains, making them, and coming back to sleep in any available hotel bathtub in town.

Okay, but you have overwrought religious adjectives to describe all of this, right?

[Dean's] "criminality" was not something that sulked and sneered; it was a wild yea-saying overburst of American joy; it was Western, the west wind, an ode from the Plains, something new, long prophesied, long a-coming.

I feel like once you steal like a dozen cars in the space of a single book, you lose the right to have the word "criminality" in scare quotes.

But please, tell us more:

[Ed and Dean] had just been laid off from the railroad. Ed had met a girl called Galatea who was living in San Francisco on her savings. These two mindless cads decided to bring the girl along [on one of their seven zillion pointless cross-country trips] and have her foot the bill. Ed cajoled and pleaded; she wouldn't go unless he married her. In a whirlwind few days Ed Dunkel married Galatea, with Dean rushing around to get the necessary papers, and a few days before Christmas they rolled out of San Francisco at seventy miles per, headed for LA and the snowless southern road. In LA they picked up a sailor in a travel bureau and took him along for fifteen dollars' worth of gas... All along the way Galatea Dunkel, Ed's new wife, kept complaining that she was tired and wanted to sleep in a motel. If this kept up they'd spend all her money long before Virginia. Two nights she forced a stop and blew tens on motels. By the time they got to Tucson she was broke. Dean and Ed gave her the slip in a hotel lobby and resumed the voyage alone, with the sailor, and without a qualm.

All right, Jack, how are you gonna justify *this* one?

Dean was simply a youth tremendously excited with life, and though he was a con-man he was only conning because he wanted so much to live and to get involved with people who would otherwise pay no attention to him.

I too enjoy life. Yet somehow this has never led me to get my friend to marry a woman in order to take her life savings, then leave her stranded in a strange city five hundred miles from home after the money runs out.

Jack Kerouac's relationship with Dean can best be described as "enabler". He rarely commits any great misdeeds himself. He's just along for the ride [usually literally, generally in flagrant contravention of all applicable traffic laws] with Dean, watching him destroy people's lives, doing nothing about it, and then going into rhapsodies about how free-spirited and unencumbered and holy and mad and visionary it all is.

There's a weird tension here, because Jack is determined to totally ignore the moral issues. He brings this kind of stuff up only incidentally, as Exhibits A and B to support his case that Dean Moriarty is the freest and most perfect and most wonderful human being on Earth, and sort of moves past it before it becomes awkward. An enthusiastic reader, caught up in the spirit of the book, might easily miss it. The only place it is ever made explicit is page 185, when Galatea (who has since found her way back to San Francisco) confronts Dean about the trail of broken lives he's left behind him, saying:

You have absolutely no regard for anybody but yourself and your damned kicks. All you think about is what's hanging between your legs and how much money or fun you can get out of people and then you just throw them aside. Not only that, but you're silly about it. It never occurs to you that life is serious and there are people trying to make something decent out of it instead of just goofing all the time."

This, 185 pages in, is the first and last time anyone seriously tries to criticize Dean. Dean has stolen about a dozen cars. He has married one woman, had an affair with another, played the two of them off against each other, divorced the first, married the second, deserted the second with a young child whom she has no money to support, gone back to the first, dumped the first again so suddenly she has to become a prostitute to make ends meet. Later he will go back to the second, beat the first so hard that he injures his thumb and has to get it amputated, break into the second's house with a gun to kill her but change his mind, desert the second again also with a child whom she has no money to support, start dating a third, desert the third *also* with a child whom she has no money to support, and go back to the second, all while having like twenty or thirty lesser affairs on the side. As quoted above, he dumped poor Galatea in

Tucson, and later he will dump Jack in Mexico because Jack has gotten deathly ill and this is cramping his style.

So Galatea's complaint is not exactly coming out of thin air.

Jack, someone has just accused your man-crush of being selfish and goofing off all the time. Care to defend him with overwrought religious adjectives?

That's what Dean was, the HOLY GOOF... he was BEAT, the root, the soul of beatific. What was he knowing? He tried all in his power to tell me what he was knowing, and they envied that about me, my position at his side, defending him and drinking him in as they once tried to do

Right. That's the problem. People are just *jealous*, because holy ecstatic angelic Dean Moriarty likes you more than he likes them. Get a life.

II

But of course getting a life – in the sense of a home, a stable relationship, a steady job, et cetera – is exactly what all the characters in *On The Road* are desperately trying to avoid.

"Beat" has many meanings, but one of them is supposed to be "beaten down". The characters consider themselves oppressed, on the receiving end of a system that grinds them up and spits them out. This is productively compared with their total lack of any actual oppression whatsoever.

I don't know if it's the time period or merely their personal charm, but Kerouac et al's ability to do anything (and anyone) and get away with it is astounding. Several of their titular cross-country trips are performed entirely by hitch-hiking, with their drivers often willing to buy them food along the way. Another is performed in some sort of incredibly ritzy Cadillac limo, because a rich man wants his Cadillac transported from Denver to Chicago, Dean volunteers, and the rich man moronically accepts. Dean of course starts driving at 110 mph, gets in an accident, and ends up with the car half destroyed. Once in the city, Dean decides this is a good way to pick up girls, and:

In his mad frenzy Dean backed up smack on hydrants and tittered maniacally. By nine o' clock the car was an utter wreck: the brakes weren't working anymore; the fenders were stove in; the rods were rattling. Dean couldn't stop it at red lights; it kept kicking convulsively over the roadway. It had paid the price of the night. It was a muddy boot and no longer a shiny limousine... 'Whee!' It was now time to return the Cadillac to the owner, who lived out on Lake Shore Drive in a swank apartment with an enormous garage underneath managed by oil-scarred Negroes. The mechanic did not recognize the Cadillac. We handed the papers over. He scratched his head at the sight of it. We had to get out fast. We did. We took a bus back to downtown Chicago and that was that. And we never heard a word from our Chicago baron about the condition of his car, in spite of the fact that he had our addresses and could have complained.

Even more interesting than their ease of transportation to me was their ease at getting jobs. This is so obvious to them it is left unspoken. Whenever their money runs out, be they in Truckee or Texas or Toledo, they just hop over to the nearest farm or factory or whatever, say "Job, please!" and are earning back their depleted savings in no time. This is really the crux of their way of life. They don't feel bound to any one place, because traveling isn't really a risk. Be it for a week or six months, there's always going to be work waiting for them when they need it. It doesn't matter that Dean has no college degree, or a criminal history a mile long, or is only going to be in town a couple of weeks. This just seems to be a background assumption. It is most obvious when it is violated; the times it takes an entire week to find a

job, and they are complaining bitterly. Or the time the only jobs available are backbreaking farm labor, and so Jack moves on (of course abandoning the girl he is with at the time) to greener pastures that he knows are waiting.

Even more interesting than their ease of employment is their ease with women. This is unintentionally a feminist novel, in that once you read it (at least from a modern perspective) you end up realizing the vast cultural shift that had to (has to?) take place in order to protect women from people like the authors. Poor Galatea Dunkel seems to have been more of the rule than the exception – go find a pretty girl, tell her you love her, deflower her, then steal a car and drive off to do it to someone else, leaving her unmarriageable and maybe with a kid to support. Then the next time you’re back in town, look her up, give her a fake apology in order to calm her down enough for her to be willing to have sex with you again, and repeat the entire process. Here is a typical encounter with a pretty girl:

Not five nights later we went to a party in New York and I saw a girl called Inez and told her I had a friend with me that she ought to meet sometime. I was drunk and told her he was a cowboy. “Oh, I’ve always wanted to meet a cowboy.”

“Dean?” I yelled across the party. “Come over here, man!” Dean came bashfully over. An hour later, in the drunkenness and chiciness of the party, he was kneeling on the floor with his chin on her belly and telling her and promising her everything and sweating. She was a big, sexy brunette – as Garcia said, something straight out of Degas, and generally like a beautiful Parisian coquette. In a matter of days they were dickering with Camille in San Francisco by long-distance telephone for the necessary divorce papers so they could get married. Not only that, but a few months later Camille gave birth to Dean’s second baby, the result of a few nights’ rapport early in the year. And another matter of months and Inez had a baby. With one illegitimate child on the West somewhere, Dean then had four little ones, and not a cent, and was all troubles and ecstasy and speed as ever.

In case you’re wondering, Dean then runs off to Mexico, leaves Inez behind, screws a bunch of Mexican women, and eventually gets back with Camille, who is happy to have him. Seriously, if I had read this book when I was writing [Radicalizing The Romanceless](#), Dean (and his friends) would have been right up there with Henry as Exhibit B. The only punishment he ever gets for his misadventures is hitting one girlfriend in the face so hard that he breaks his own thumb, which gets infected and has to be amputated. Human justice has failed so miserably, one feels, that God has to personally step in.

As bad as the gender stuff is, the race stuff is worse. This is 1950-something, so I’m prepared for a lot of awful stuff regarding race. But this is totally *different* awful stuff regarding race than I expected. I have never been able to get upset over “exoticization” and “Orientalism” before, but this book reached new lows for me:

At lilac evening I walked with every muscle aching among the lights of 27th and Welton in the Denver colored section, wishing I were a Negro, feeling that the best the white world had offered was not enough ecstasy for me, not enough life, joy, kicks, darkness, music, not enough night. I stopped at a little shack where a man sold hot red chili in paper containers; I bought some and ate it, strolling in the dark mysterious streets. I wished I were a Denver Mexican, or even a poor overworked Jap, anything but what I was so drearily, a “white man” disillusioned. All my life I’d had white ambitions; that was why I’d abandoned a good woman like Terry in the San Joaquin Valley... a gang of colored women came by, and one of the young ones detached herself from motherlike elders and came to me fast – “Hello Joe!” and suddenly saw it wasn’t Joe, and ran back blushing. I wished I were Joe. I was only myself, sad, strolling in this violet dark, this unbearably sweet night, wishing I could exchange worlds with the happy, true-hearted, ecstatic Negroes of America.

Negroes are holy and ecstatic. But only in the same way barns and cows are holy and ecstatic. One gets the suspicion

that Jack Kerouac is not exactly interacting with any of this stuff, so much as using it as something he can have his overwrought religious feelings about.

The “heroes” of *On The Road* consider themselves ill-done by and beaten-down. But they are people who can go anywhere they want for free, get a job any time they want, hook up with any girl in the country, and be so clueless about the world that they’re pretty sure being a 1950s black person is a laugh a minute. *On The Road* seems to be a picture of a high-trust society. Drivers assume hitchhikers are trustworthy and will take them anywhere. Women assume men are trustworthy and will accept any promise. Employers assume workers are trustworthy and don’t bother with background checks. It’s pretty neat.

But *On The Road* is, most importantly, a picture of a high-trust society collapsing. And it’s collapsing precisely because the book’s protagonists are going around defecting against everyone they meet at a hundred ten miles an hour.

III

The viewpoint of a character in a book is not necessarily the viewpoint of its author. One can write about terrible people doing terrible things and not necessarily endorse it. That having been said, it’s very hard to read Jack Kerouac-the-author as differing very much from Jack-Kerouac-the-character in his opinions. He still has a raging man-crush on Dean and thinks that he is some kind of holy madman who can do no wrong.

The nicest thing I can say about *On The Road* is that perhaps it should be [read backwards](#). It is a paean to a life made without compromise, a life of enjoying the hidden beauty of the world, spent in pursuit of holiness and the exotic. Despite how I probably sound, I really respect the Beat aesthetic of searching for transcendence and finding it everywhere. There’s something to be said for living your life to maximize that kind of thing, especially if everyone else is some kind of boring dispirited factory worker or something. Kerouac wrote around the same time as Sartre; it’s not difficult to imagine him as one of the first people saying you needed to try to find your True Self.

Read backwards, there was a time when to spend your twenties traveling the world and sleeping with strange women and having faux mystical experiences was something new and exciting and dangerous and for all anybody knew maybe it held the secret to immense spiritual growth. But from a modern perspective, if Jack and Dean tried the same thing today, they’d be one of about a billion college students and aimless twenty-somethings with exactly the same idea, posting their photos to Instagram tagged “holy”, “ecstatic”, and “angelic”. There’s nothing wrong with that. But it doesn’t seem like a good stopping-point for a philosophy. It doesn’t even seem like good escapism. I’d be willing to tolerate all the pointless criminality if it spoke to the secret things that I’ve always wanted to do in my hidden heart of hearts, but I’d like to think there’s more there than driving back and forth and going to what seem like kind of lackluster parties.

When I [read Marx](#), I thought that his key mistake was a negative view of utopia. That is, utopia is what happens automatically once you overthrow all of the people and structures who are preventing there from being utopia. Just get rid of the capitalists, and the World-Spirit will take care of the rest. The thought that ordinary, fallible, non-World-Spirit humans will have to build the post-revolution world brick by brick, and there’s no guarantee they will do any better than the pre-revolutionary humans who did the same, never seems to have occurred to him.

Kerouac was a staunch anti-Communist, but his beat philosophy seems to share the same wellspring. Once you get rid of all the shackles of society in your personal life – once you stop caring about all those squares who want you to have families and homes and careers and non-terrible friends – once you become a holy criminal who isn’t bound by the law or other people’s needs – then you’ll end up with some ecstatic visionary true self. Kerouac claimed he was Catholic, that he was in search of the Catholic God, and that he found Him – but all of his descriptions of such tend to be a couple of minutes of rapture upon seeing some especially pretty woman in a nightclub or some especially dingy

San Francisco alley, followed by continuing to be a jerk who feels driven to travel across the country approximately seven zillion times for no reason.

Like the early Communists, who were always playing up every new factory that opened as the herald of the new age of plenty, in the beginning it's easy to tell yourself your revolution is succeeding, that you are right on the brink of the new age. But at last come the Andropovs and Brezhnevs of the soul, the stagnation and despair and the going through the motions.

Kerouac apparently got married and divorced a couple of times, became an alcoholic, had a bit of a breakdown, and drank himself to death at age 49. Moriarty spent a while in prison on sort-of-trumped-up drug charges, went through a nasty divorce with whichever wife hadn't divorced him already, and died of a likely drug overdose at age 47.

Overall I did not like this book.

If you're writing about a crime spree you were a part of, you ought to show at least a little self-awareness.

Mysticism continues to be a perfectly valid life choice, but I continue to believe if you want to pursue it you should do it carefully and methodically, for example meditating for an hour a day and then going to regular retreats run by spiritual authorities, rather than the counterculture route of taking lots of drugs and having lots of sex and reading some books on Gnosticism and hoping some kind of enlightenment smashes into you.

Professional writing should be limited to about four overwrought religious adjectives per sentence, possibly by law.

And travel and girls are both fun, but *[doctor voice]* should be enjoyed responsibly and in moderation.

The Anti-Reactionary FAQ

Posted on October 20, 2013 by Scott Alexander



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.

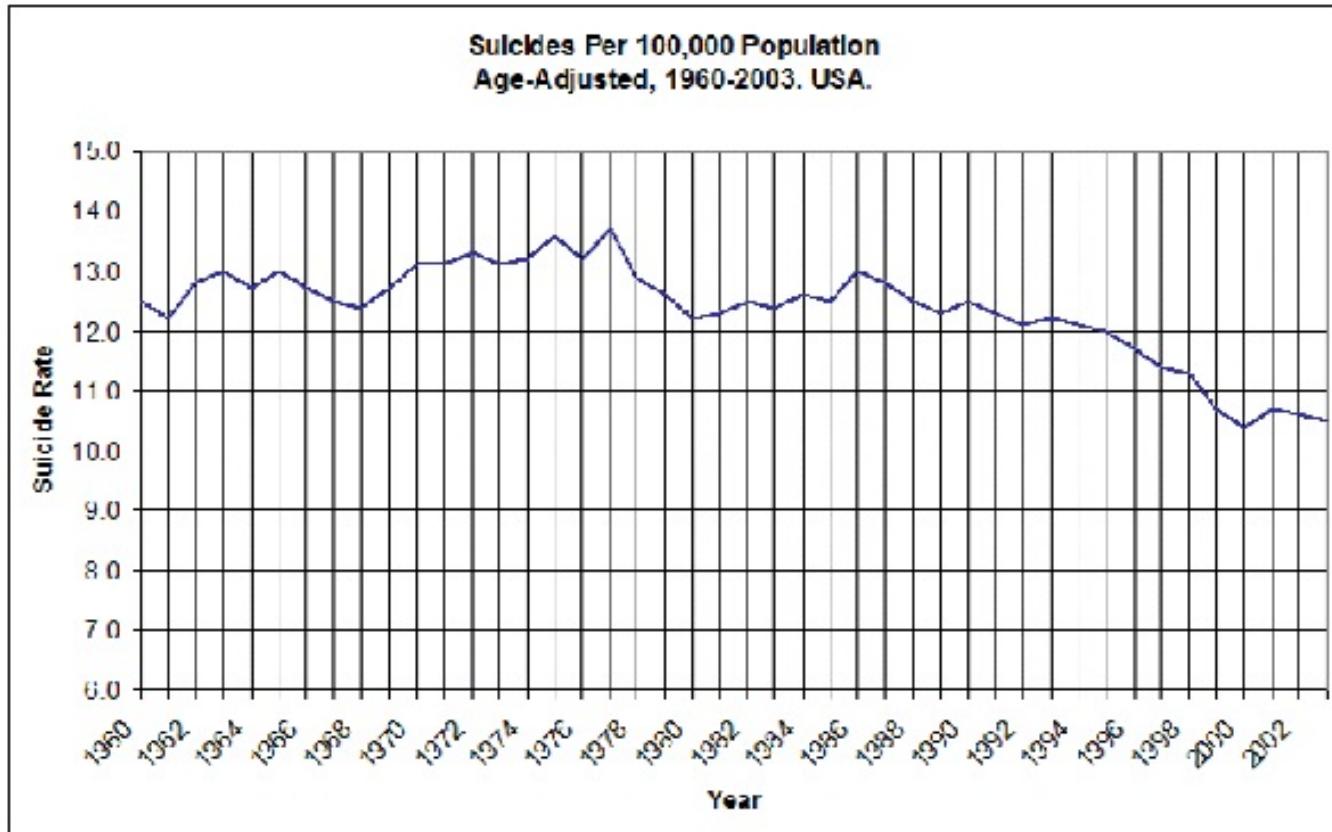
I will not require a completely monotonic downward trend, but neither will I accept a blip of one or two years in a generally positive trend as proving all modern civilization is bankrupt.

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/vitalstat.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

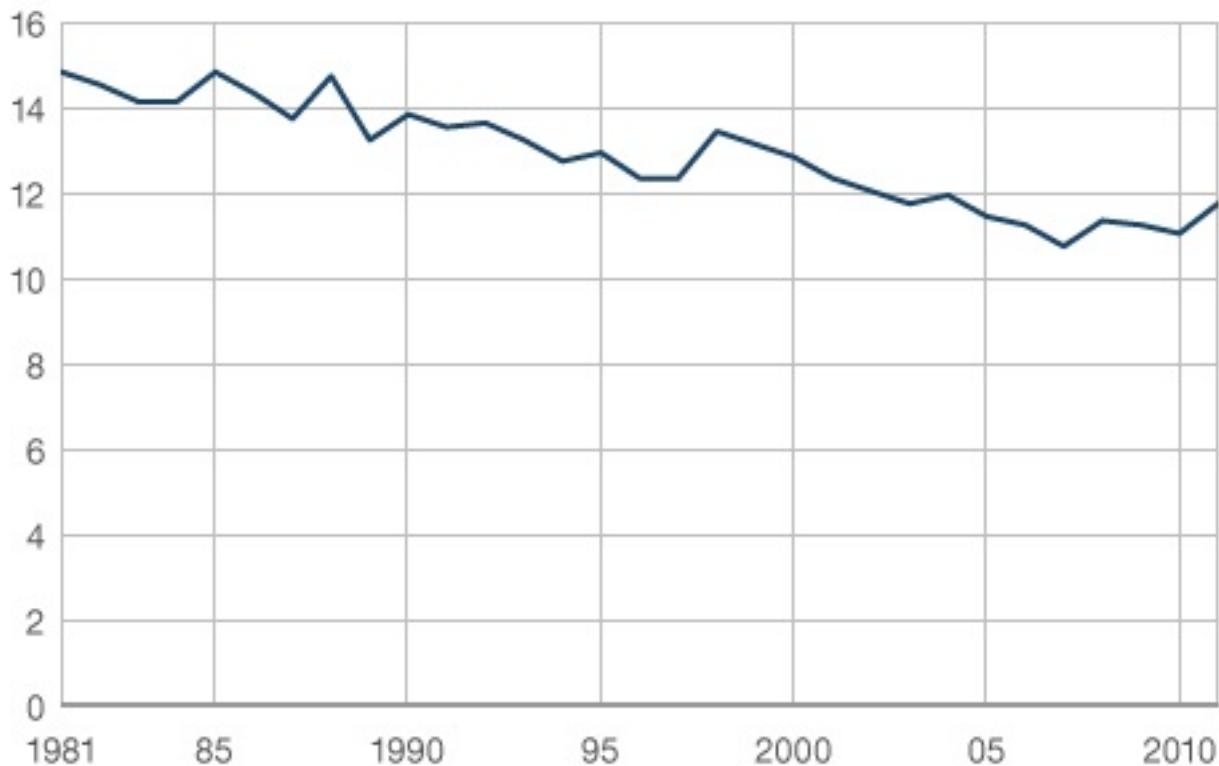
suicide rates rose from 1999 to 2010. Apparently my new job is reminding Reactionaries that they cannot blindly trust New York Times articles to give them the whole truth.

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.

Suicide rate per 100,000 people

Aged 15 and over



Source: Office for National Statistics

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.

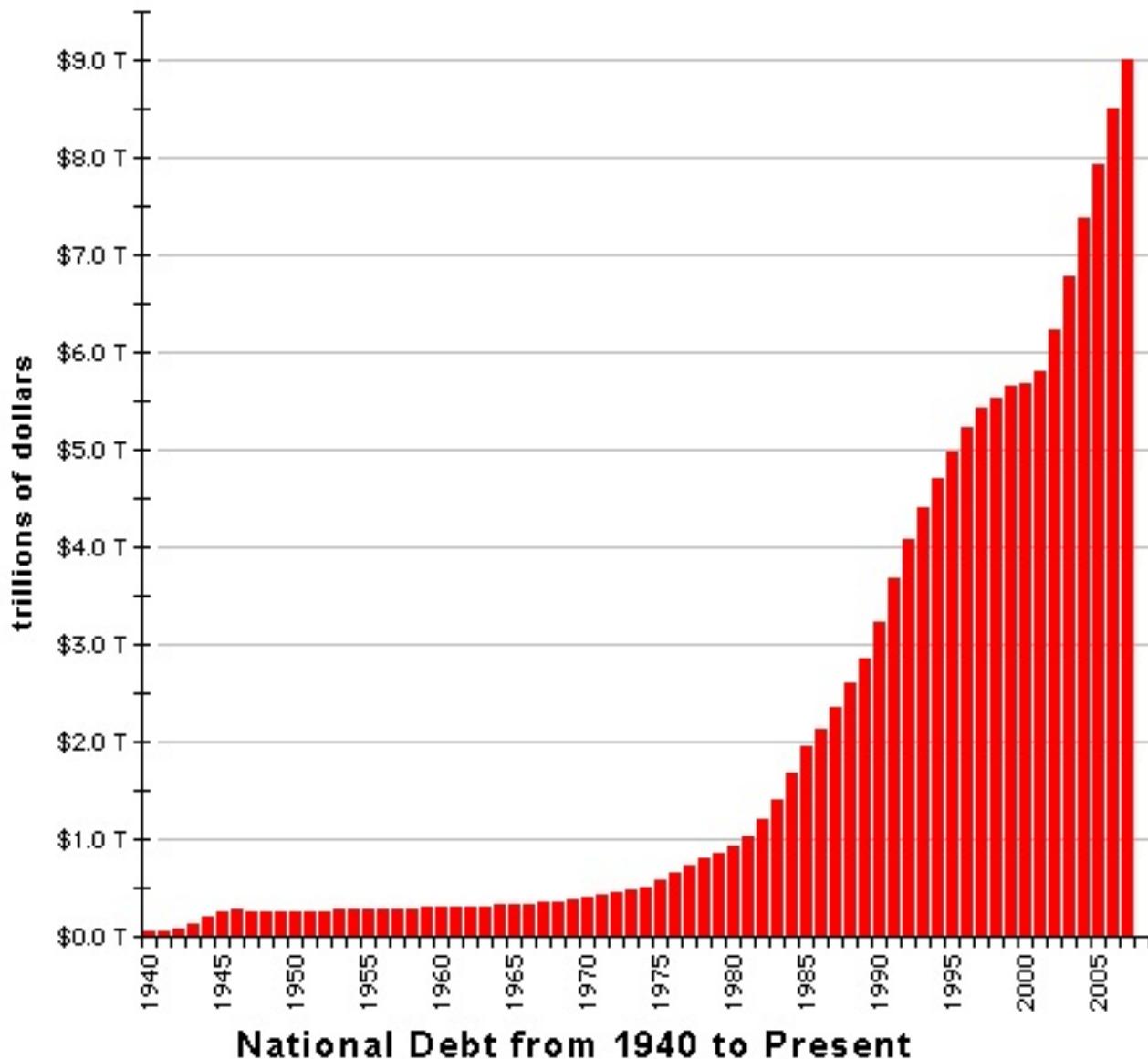
Even better, we find that suicide [peaked in Britain in 1905](#) – just after the Victorian period – and has been declining ever since.

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?

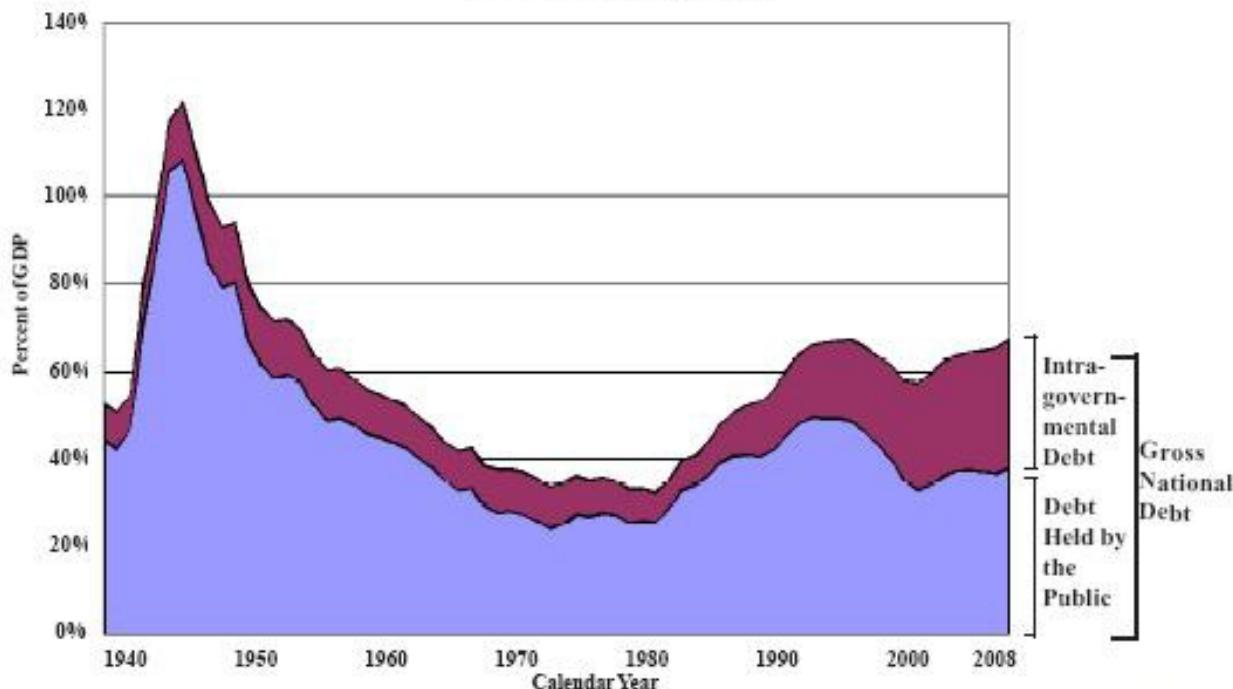
Here again the Reactionaries overstate their case. Michael tried to support his point with...



Source: U.S. National Debt Clock
http://www.brillig.com/debt_clock/

...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:

The National Debt as a Percent of GDP FY 1940-2008



Estimate for the end of 2008 -

Debt Held by the Public: \$5.4 trillion dollars (37.9% of GDP)

Intragovernmental Debt: \$4.2 trillion dollars (29.5% of GDP)

Gross National Debt : \$9.7 trillion dollars (67.5% of GDP)

Source: President's Budget, FY2008

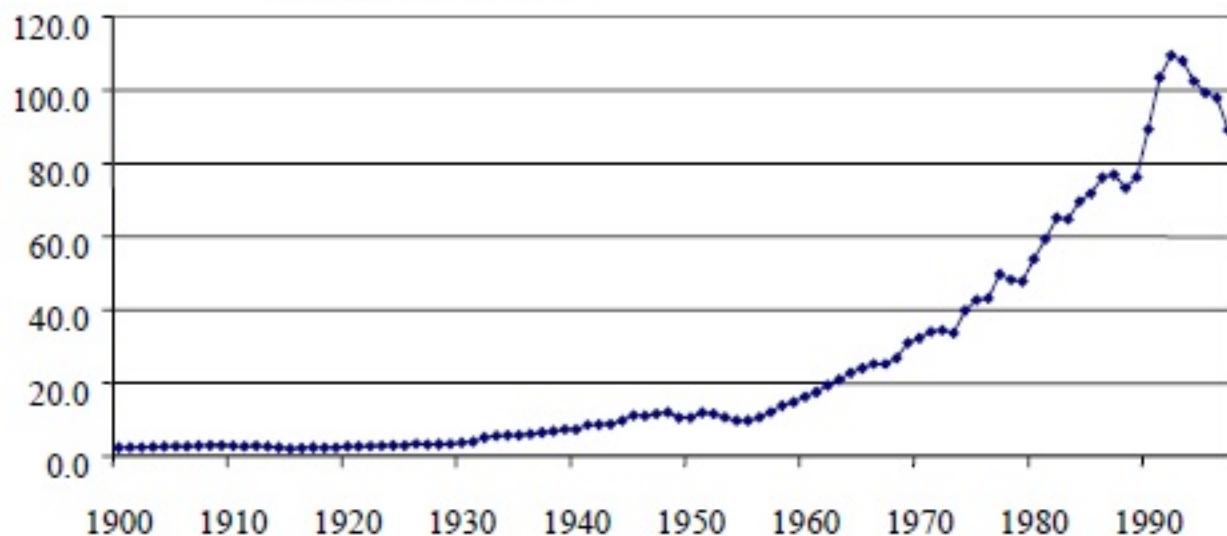


To his credit, Michael fixed this when I pointed it out. But to me, the new graph looks like gradual decrease in debt since World War II up until Reagan's big military buildup, followed by a gradual retreat from that military buildup. My God, won't somebody stop Progressivism before it's too late?!?!

1.3. Is crime becoming worse?

Michael's statistics for crime deserve more attention:

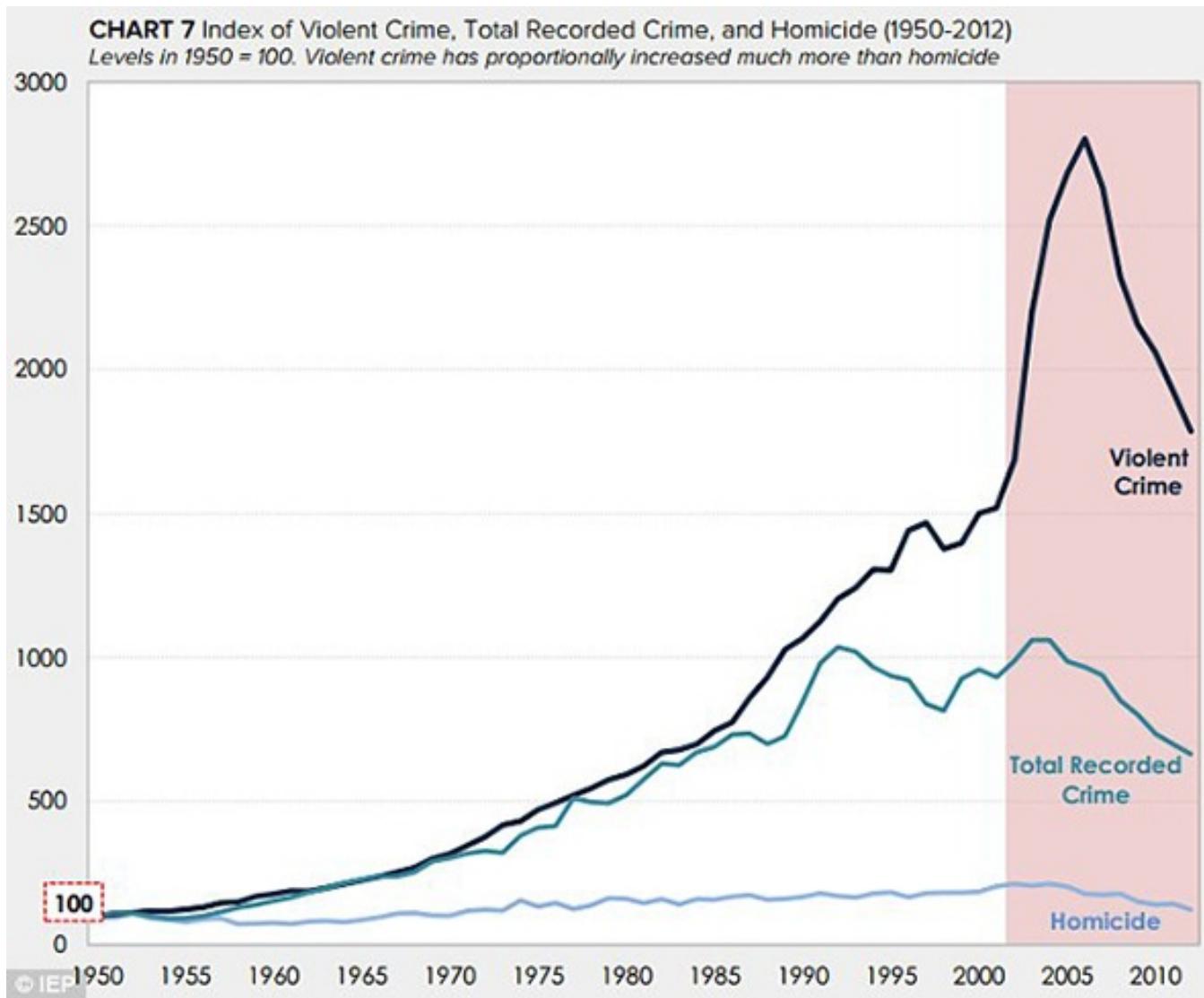
Indictable Offences Known to the Police (per thousand of population) in England & Wales 1900-1997.



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:

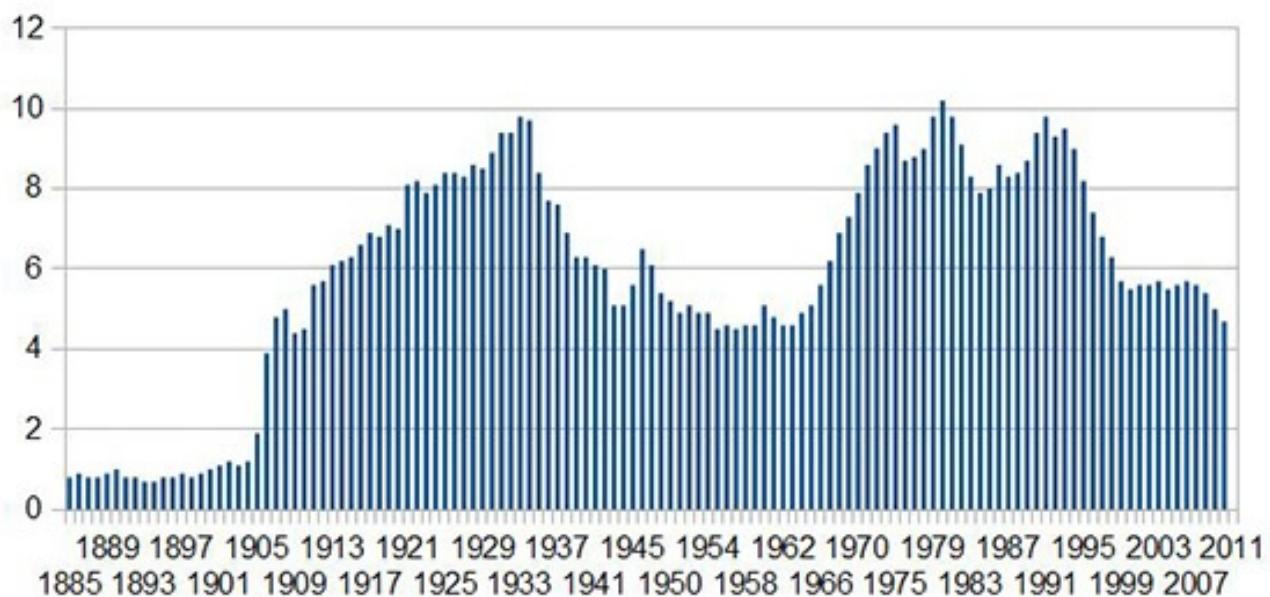
U.S. Violent Crime Rate, U.S. Justice Department Statistics, 1973-2010

Number of victims per 1,000 population aged 12 or older



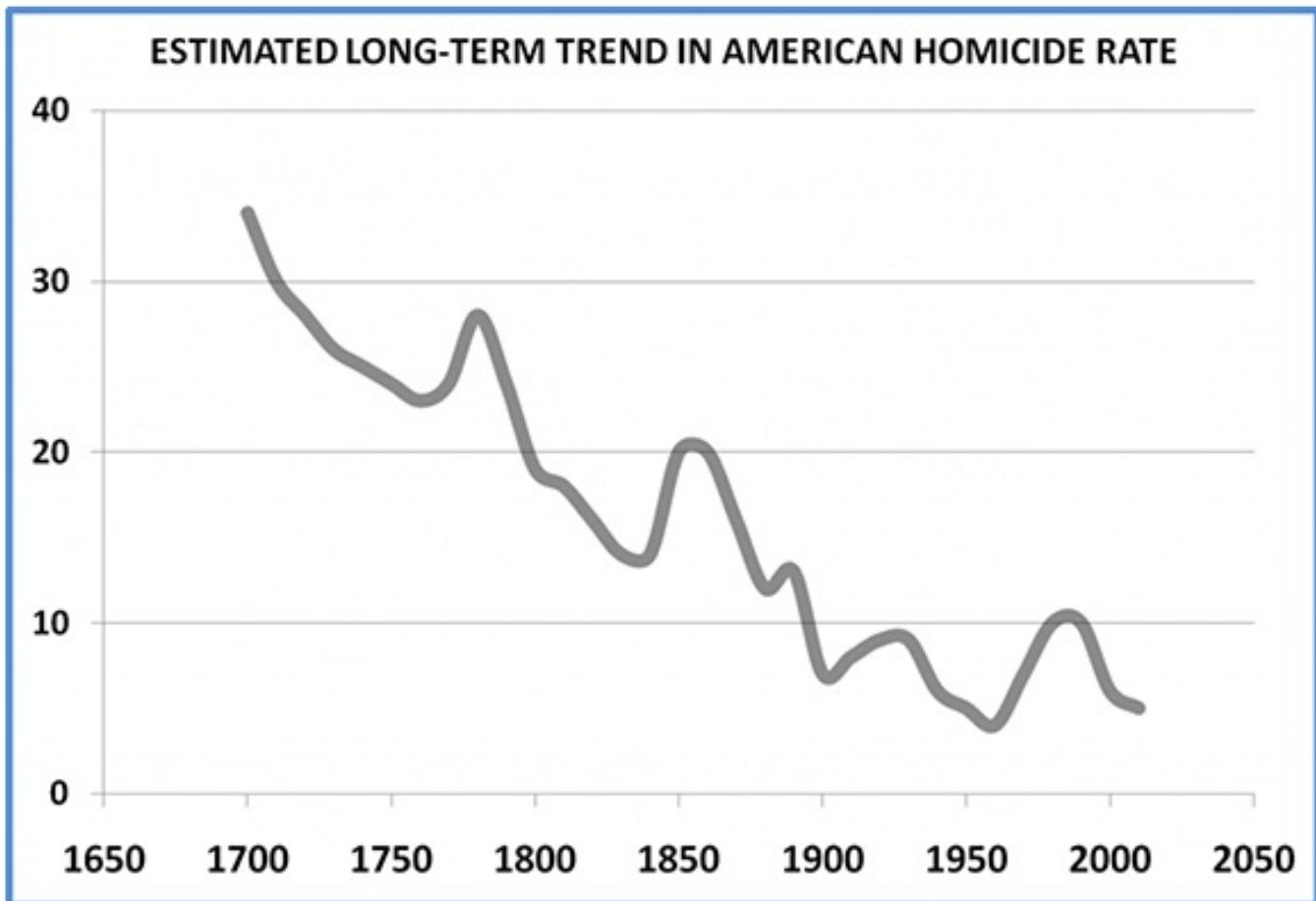
I wish I could find longer-term US crime rate data, but it doesn't seem to be out there. I can, however, find longer-term homicide data:

U.S. Homicide Rate, 1885-2010

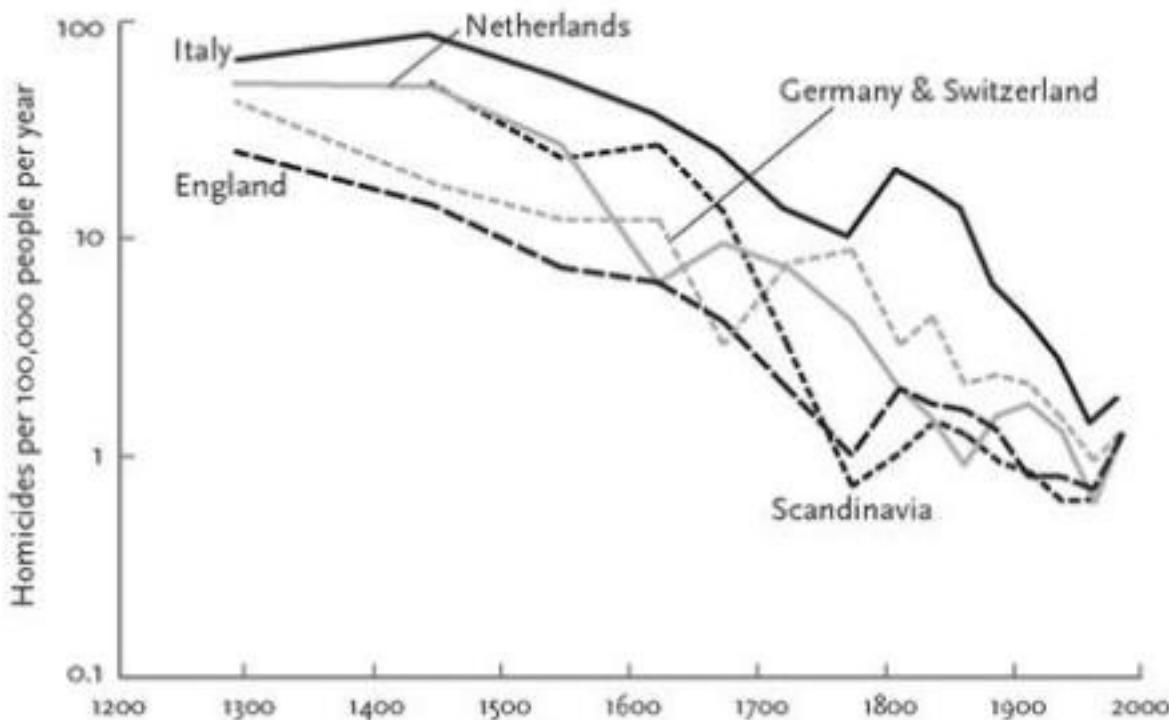


We see ups and downs but no general pattern. A Reactionary might cite the apparently very low level of homicides in 1885, but historians pretty much agree that's a reporting artifact and that [the period ending in 1887 had the highest murder rate in American history](#). In any case, right now we seem 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 was 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 garroting occurred in various parts of the country in the period from 1856 to 1862. Garroting 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 lynchpin 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.

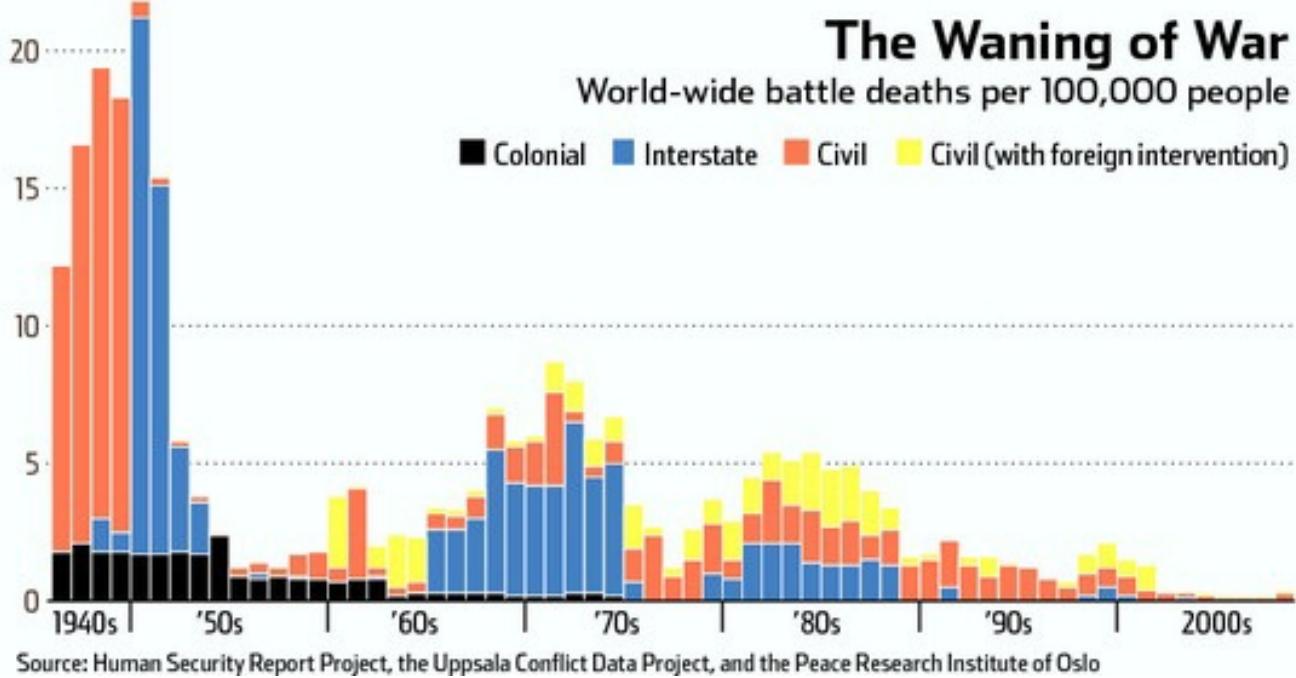
Even so, a word to defend technology. Right now I am typing a lengthy essay that will be read by a few thousand people. A couple dozen of those will discuss it in the comments. Among those will be people with whom I've had interesting discussions, friendships, and even a couple of romantic relationships. 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 another's 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](#).

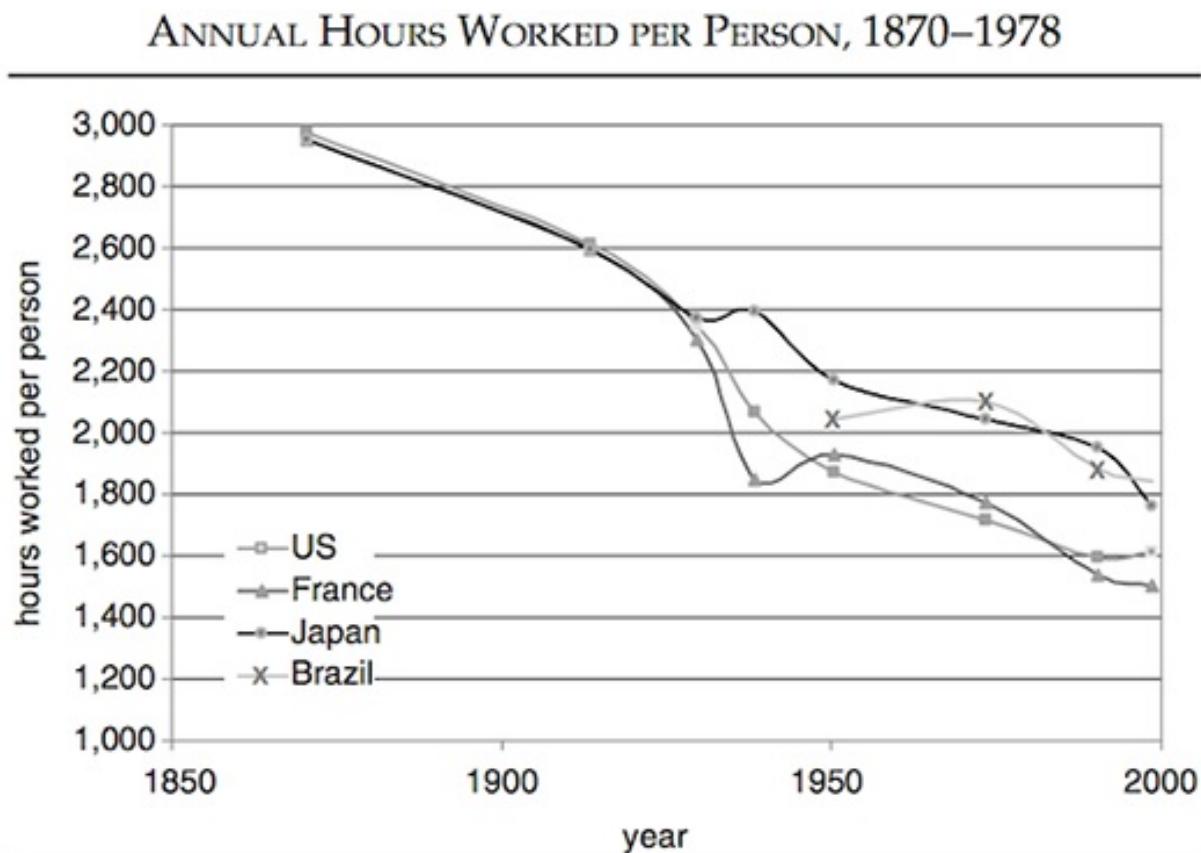
1.7.1. What about the Concert of Europe? The great statesman Klemens von Metternich used Reactionary ideas to create a brilliant system that kept peace in Europe for nearly a century!

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.

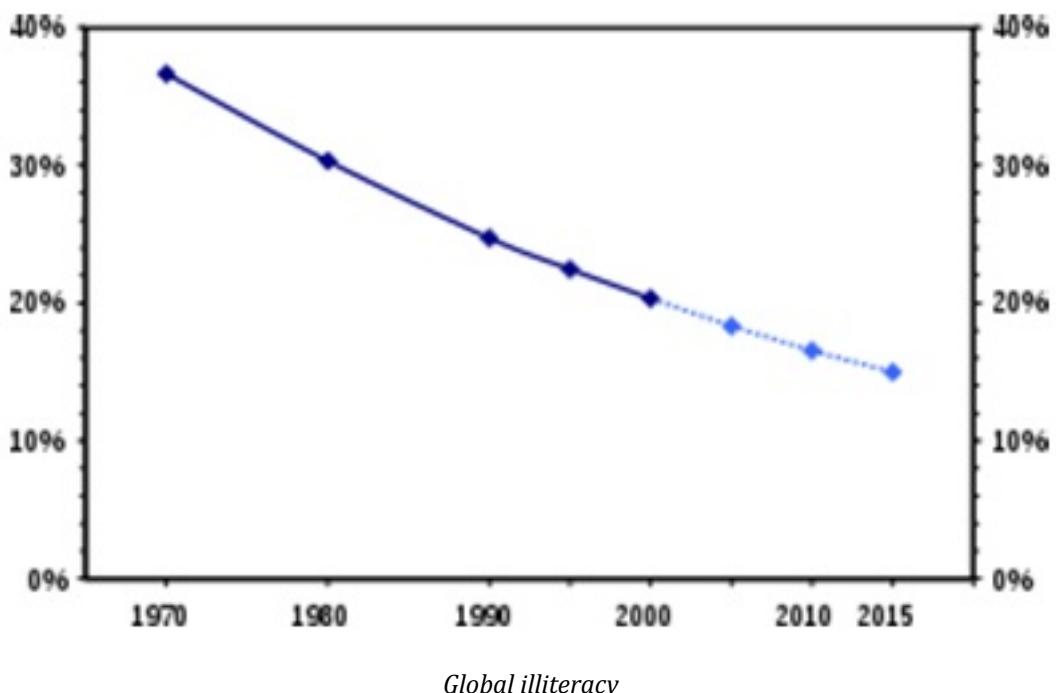
1.8. Okay, you've discussed the trends Michael listed as supporting Reaction, and found them less than convincing. Do you have any trends of your own that you think support more modern societies?

Yes. Most of the graphs below come from [31 Charts That Will Restore Your Faith In Humanity](#).

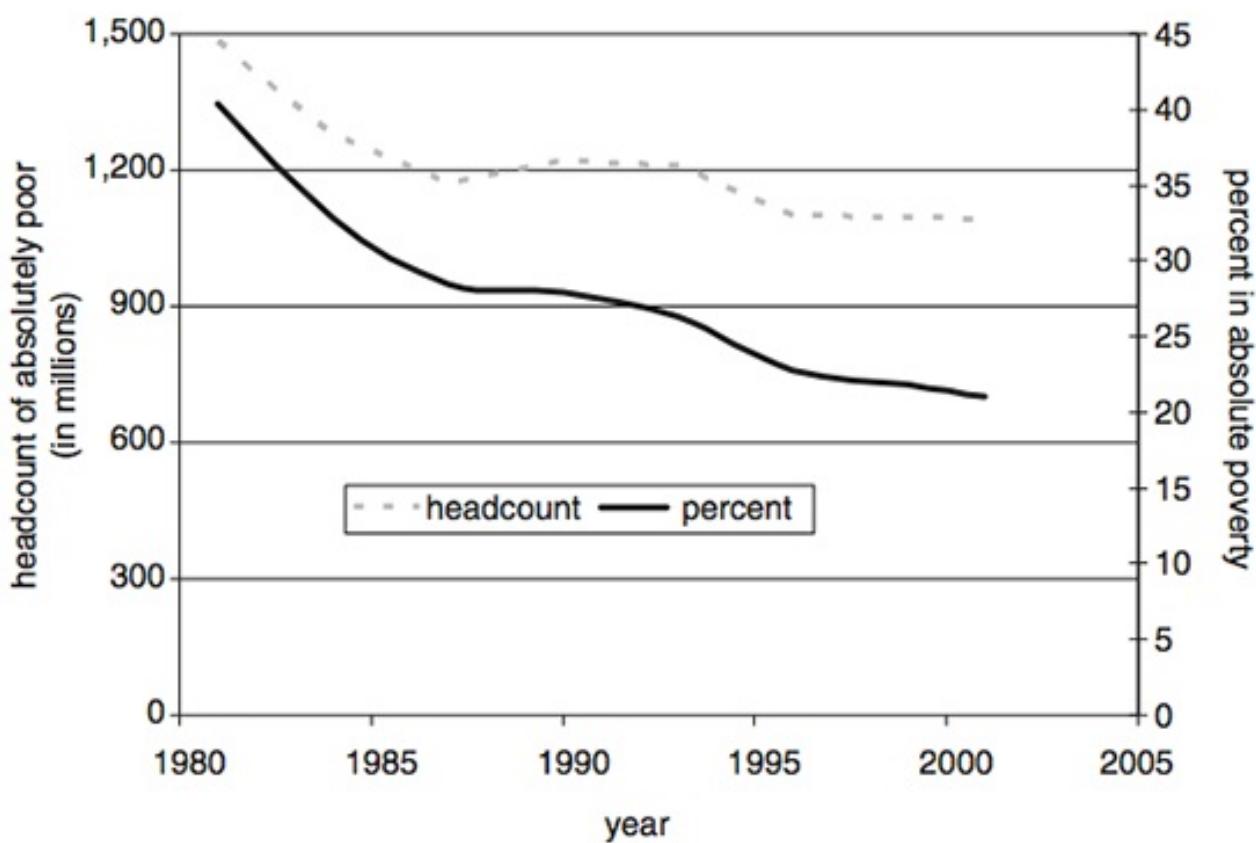


SOURCES: Angus Maddison, *Chinese Economic Performance in the Long Run* (Paris: OECD, 1998); Angus Maddison, *The World Economy: A Millennial Perspective* (Paris: OECD, 2001).

Hours worked per person

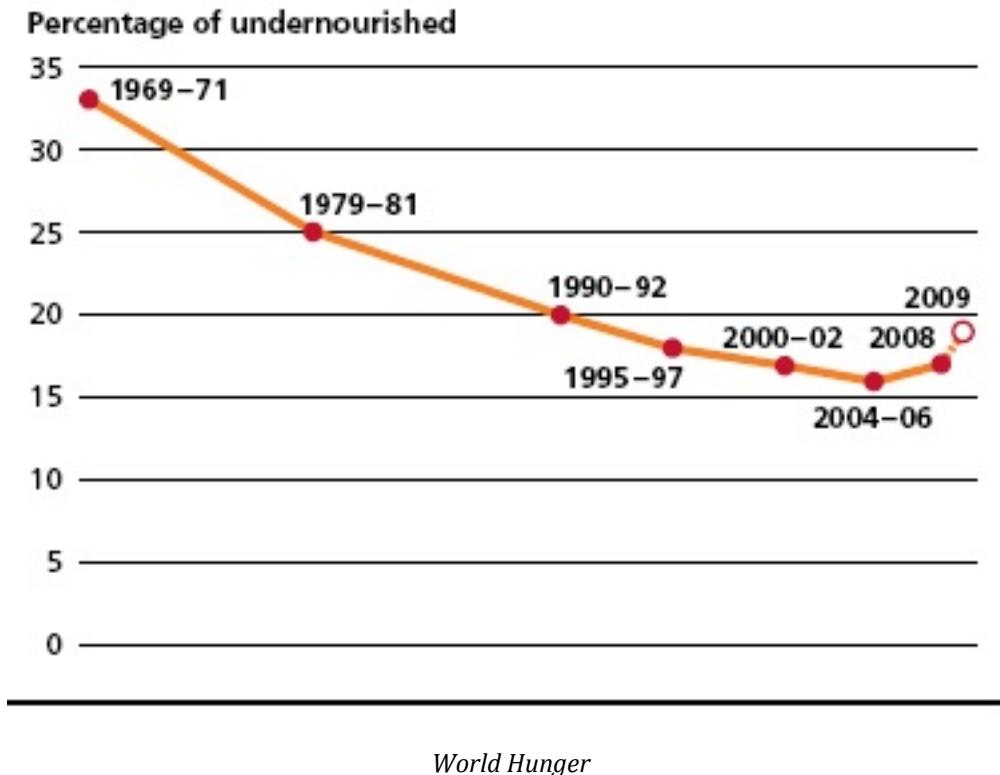


GLOBAL POVERTY, 1980–2001



SOURCE: World Bank, *World Development Indicators*, <http://devdata.worldbank.org/dataonline> (accessed July 12, 2005).

Global poverty



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 quintupling, 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 millennia?

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 Misérables*? 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 *baaaaasicallly* 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 Conqueror, 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 régime* 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.

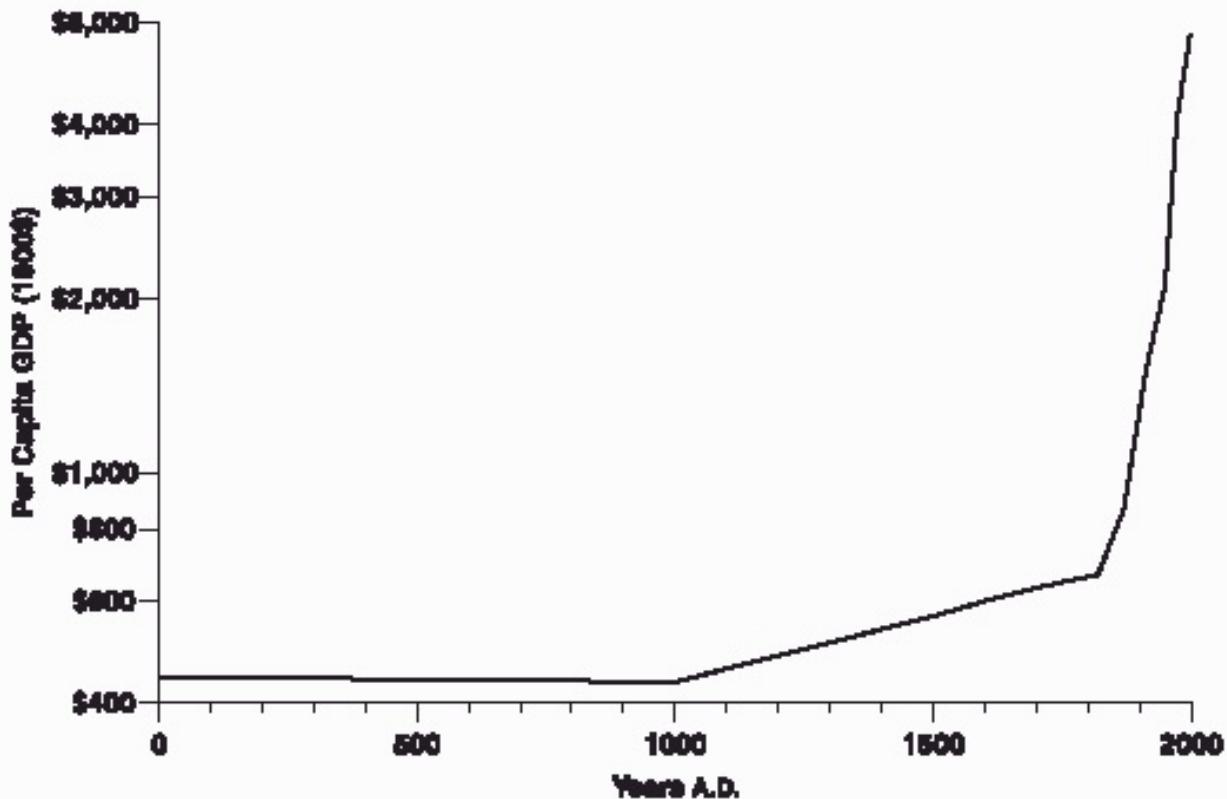
Sure, Otto von Bismarck was no hippie, but he was first and foremost a pragmatist, and his empire combined both conservative and progressive elements. It was based on a constitution, had universal male suffrage (only 5 years after the US got same!), elected a parliament, and allowed political parties. 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)



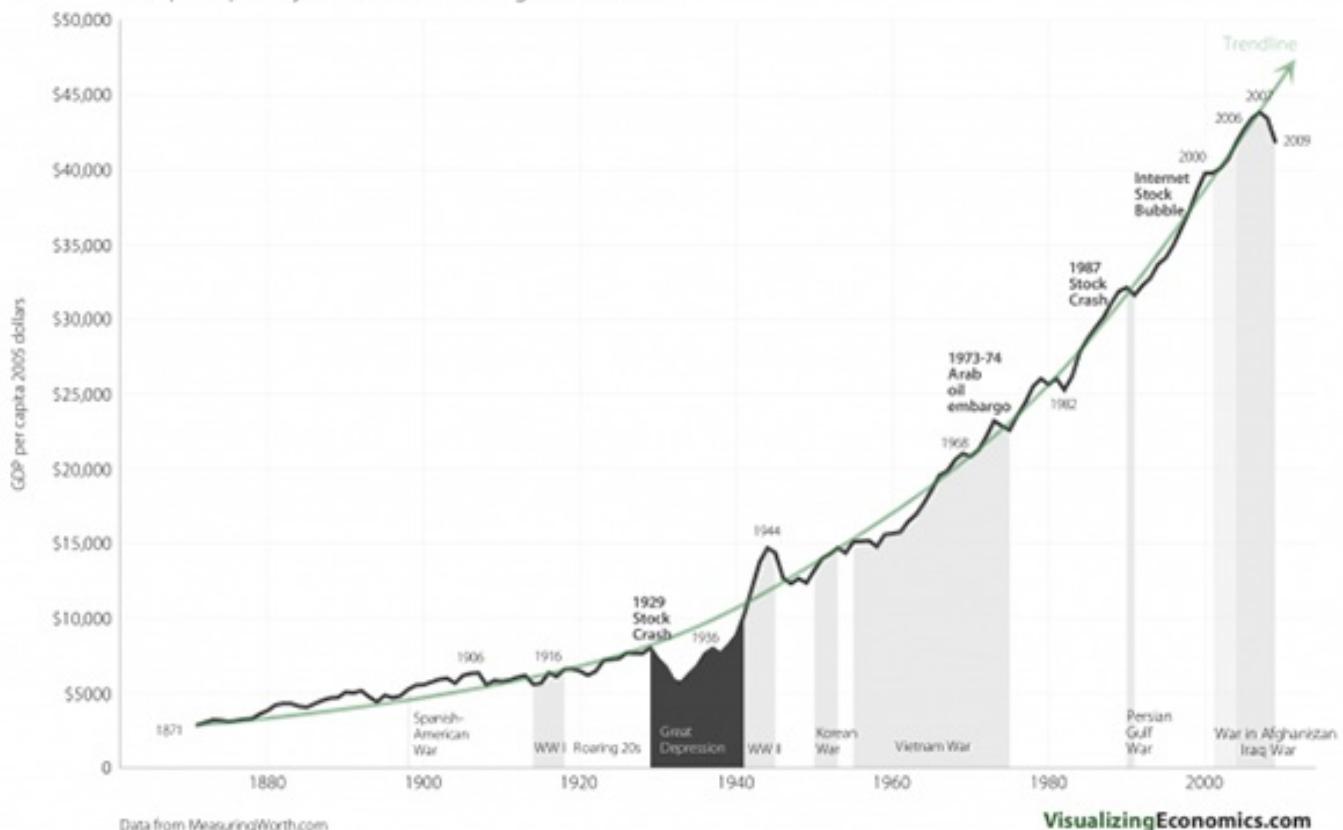
Source: Maddison, *The World Economy: A Millennial Perspective*, 264.

To put it lightly, I see no evidence of a decline starting around 1789?

Maybe the effect is just in the United States?

Long-term real growth in US GDP per capita 1871–2009

GDP per capita adjusted for inflation using 2005 dollars



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 Septimius 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 "ecumenicalism" 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 these 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 inverters 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 inverters 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 inverters. 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 LARGE 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 *somewhat* 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 year	% 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 debt	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.

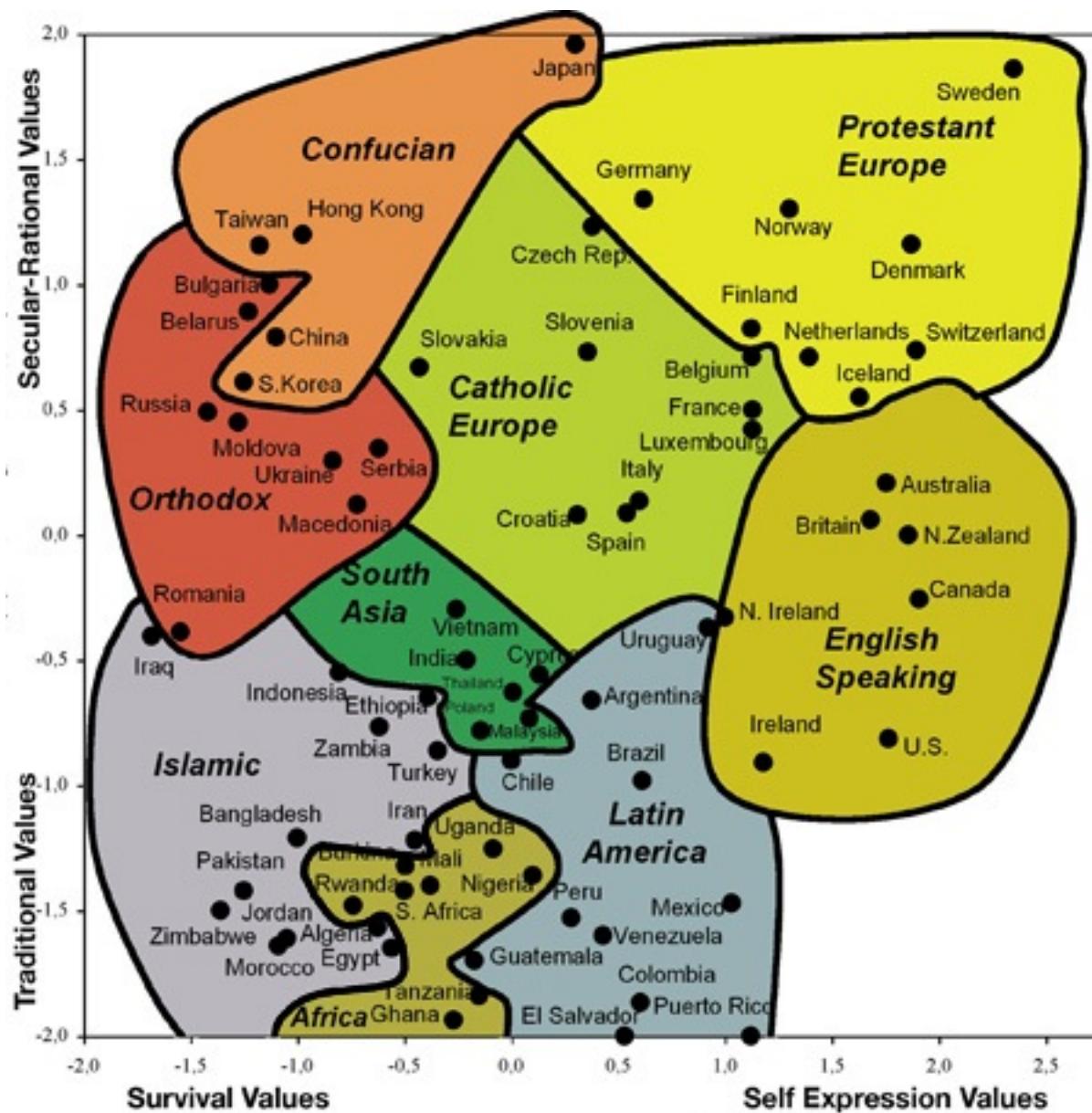
On the hottest, most politically relevant topics, society has moved leftward either very slowly or not at all. Over the past generation, it has moved to the right on gun control, the welfare state, capitalism, labor unions, and the environment. Although the particular time series on the chart does not 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?

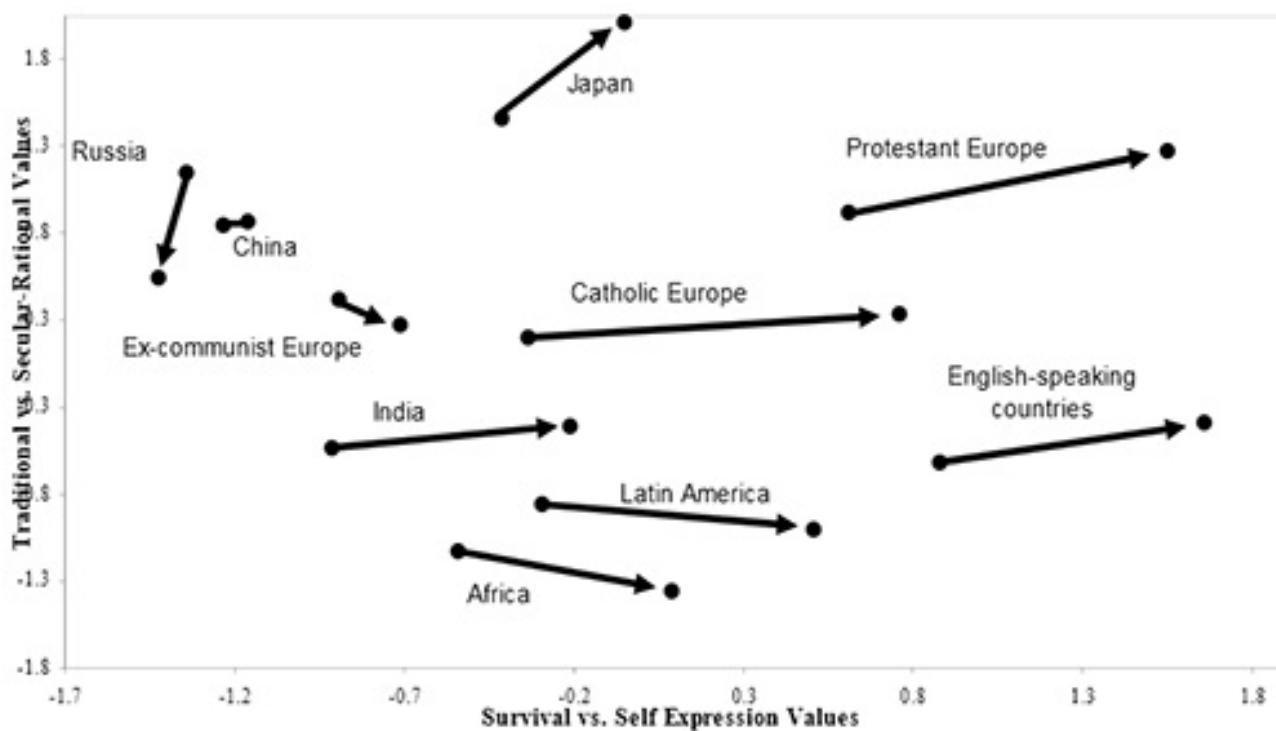
The [World Values Survey](#) is the official academic attempt to understand this question. They've been polling in eighty countries around the world for thirty years trying to figure out who has what values and how they have been changing. Maybe you've seen the most famous summary of their results:



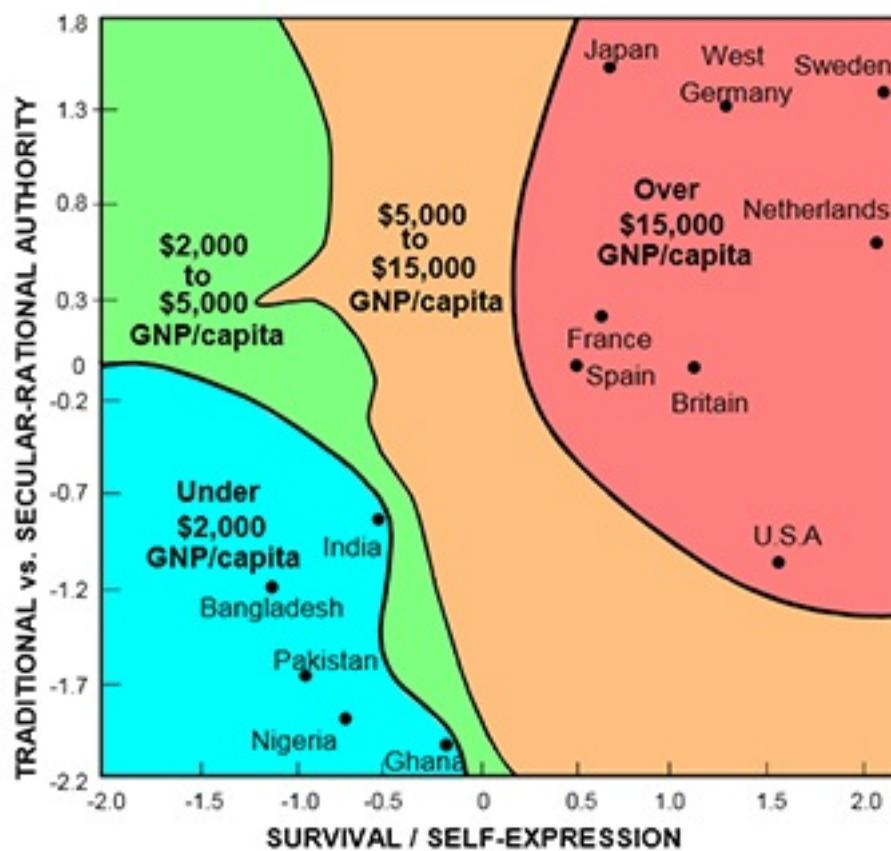
There is no end to the fun one can have with WVS data, and I highly recommend at least Wikipedia's [Catalogue of Findings](#) if not the original studies. But the most important part is that dimensionality analysis finds that answers to value questions cluster together onto two axes: survival vs. self-expression 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.

Changes over time, 1981-2007

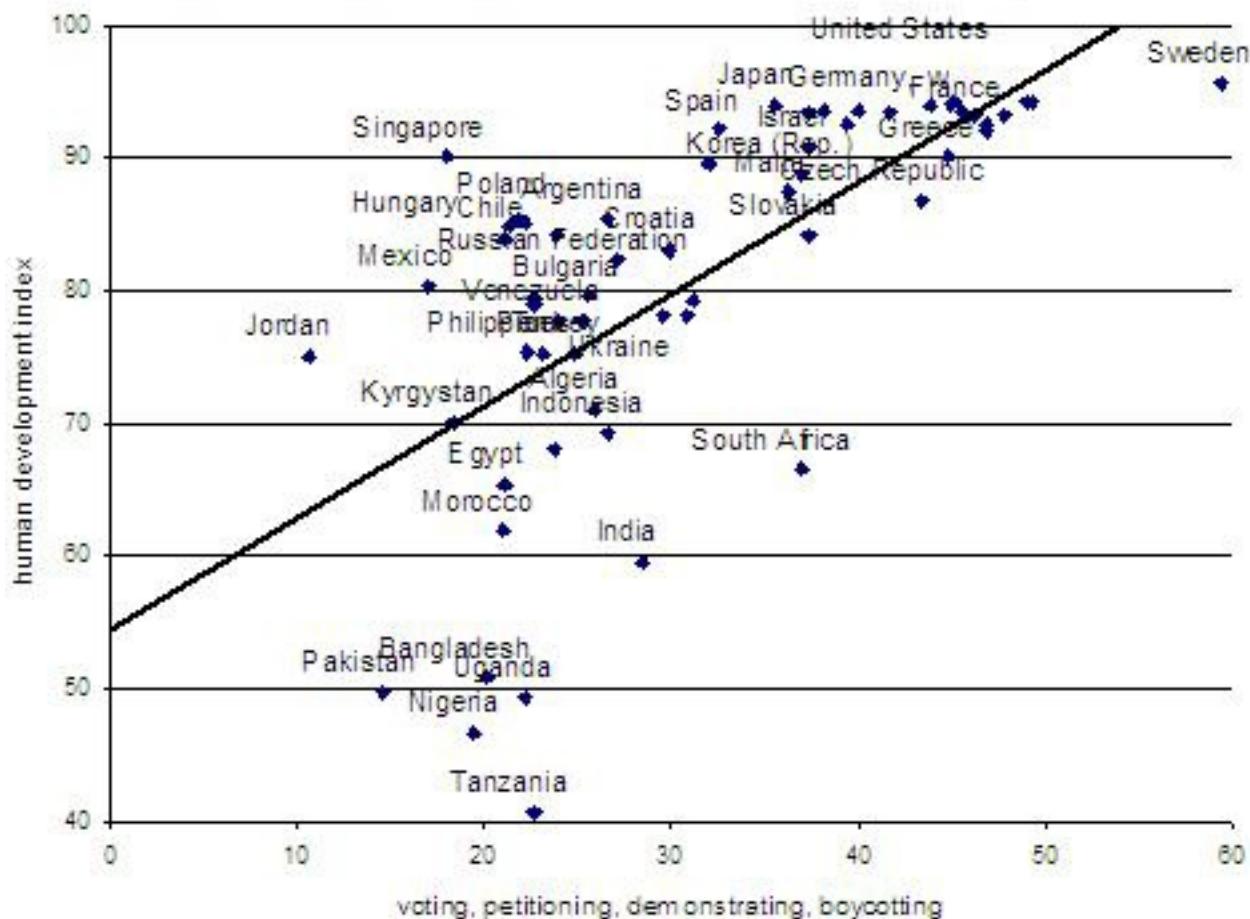


Both within a single time period and between time periods, traditional and survival values are generally associated with poverty, low industrialization, and insecurity. Secular-rational and self-expression values are generally associated with wealth, industrial or knowledge economies, and high security. The difference is not subtle:



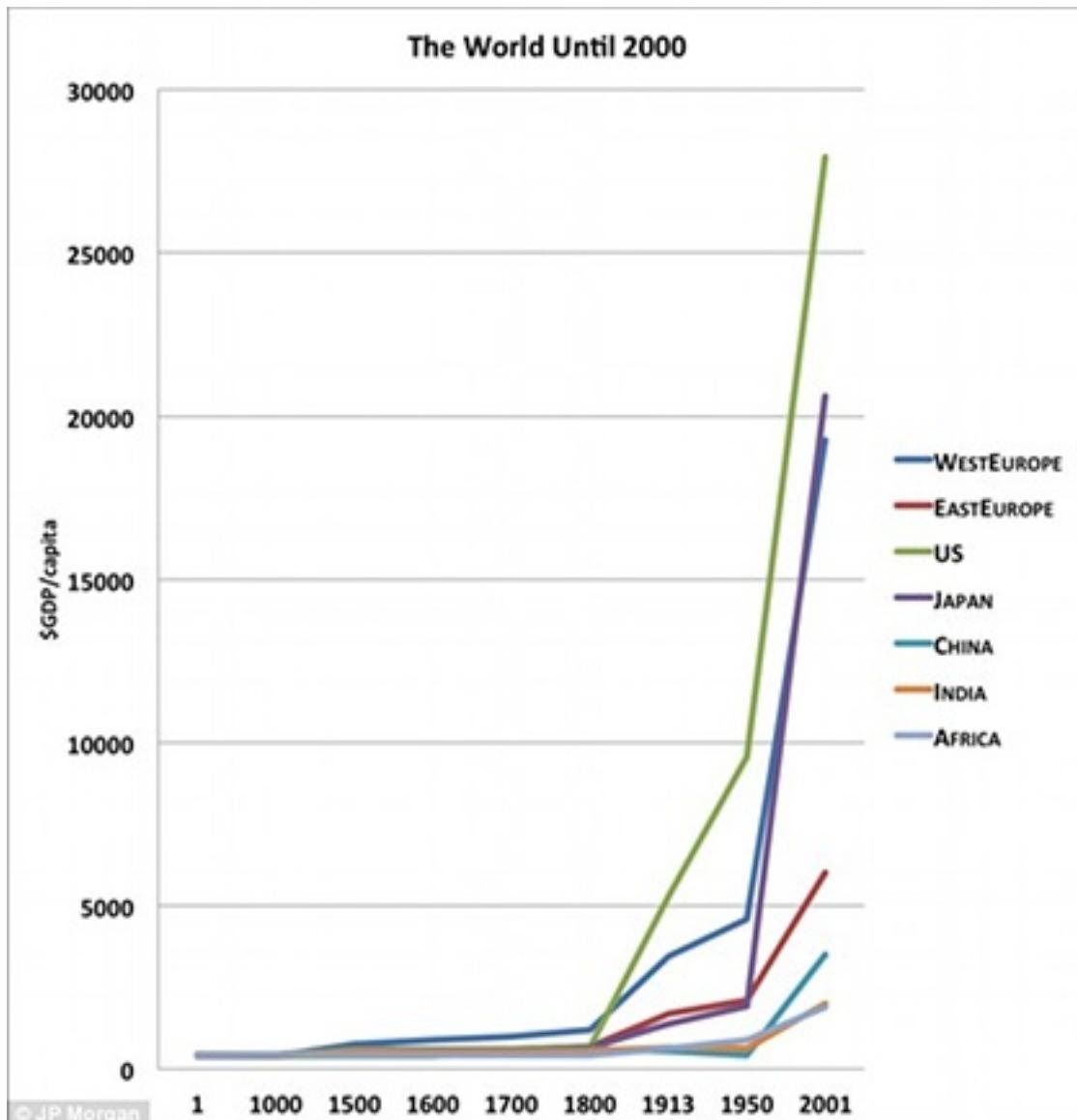
And if you want to know why countries are becoming more democratic and less monarchist, it's hard to get a more direct answer than this graph (although its attempt at a linear fit was a bad idea):

political participation and human development



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:



Riiight 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.

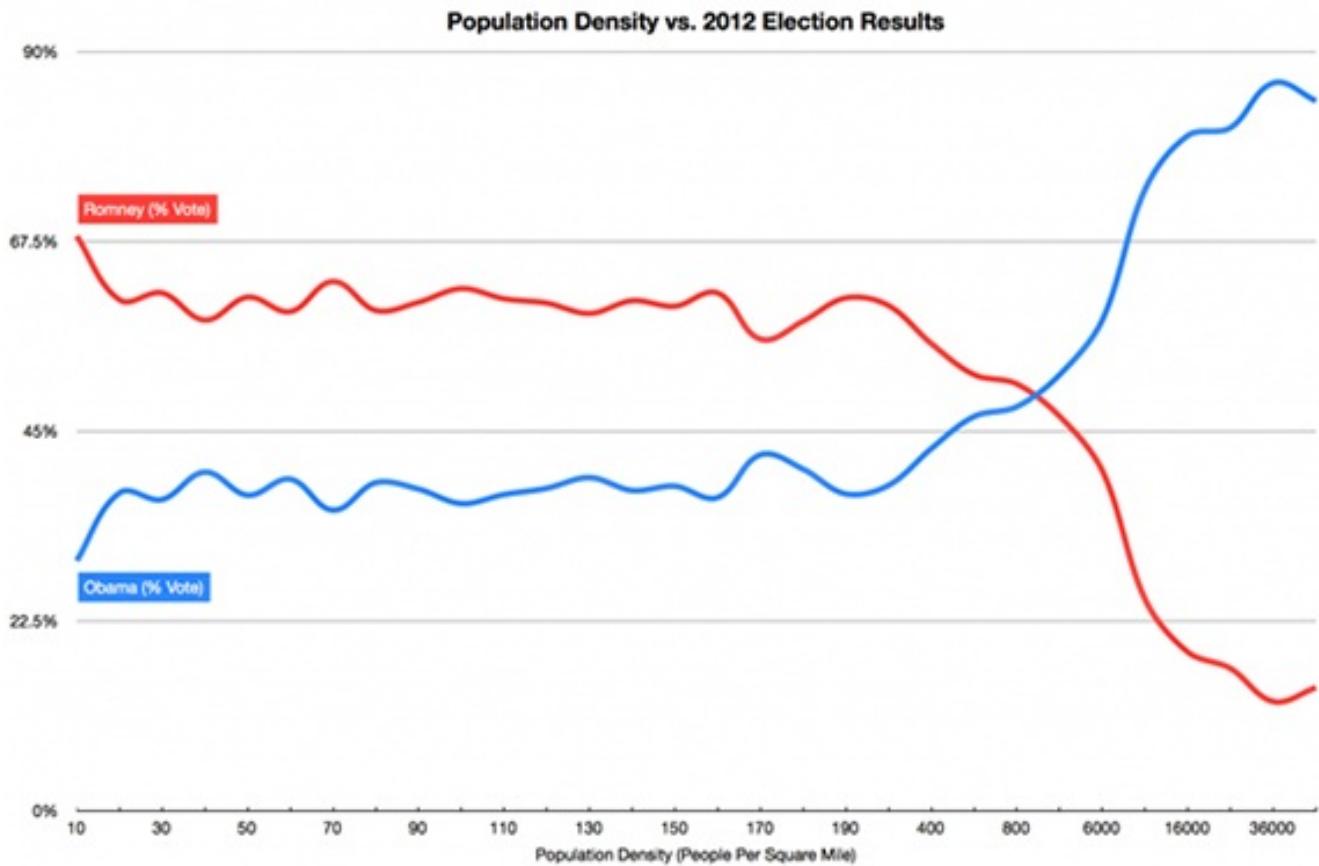
And, well, right about the time of the French Revolution Europe and the USA started shifting to the left, with Japan following after World War II, and Africa and Southeast Asia still lagging behind.

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/Survive 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

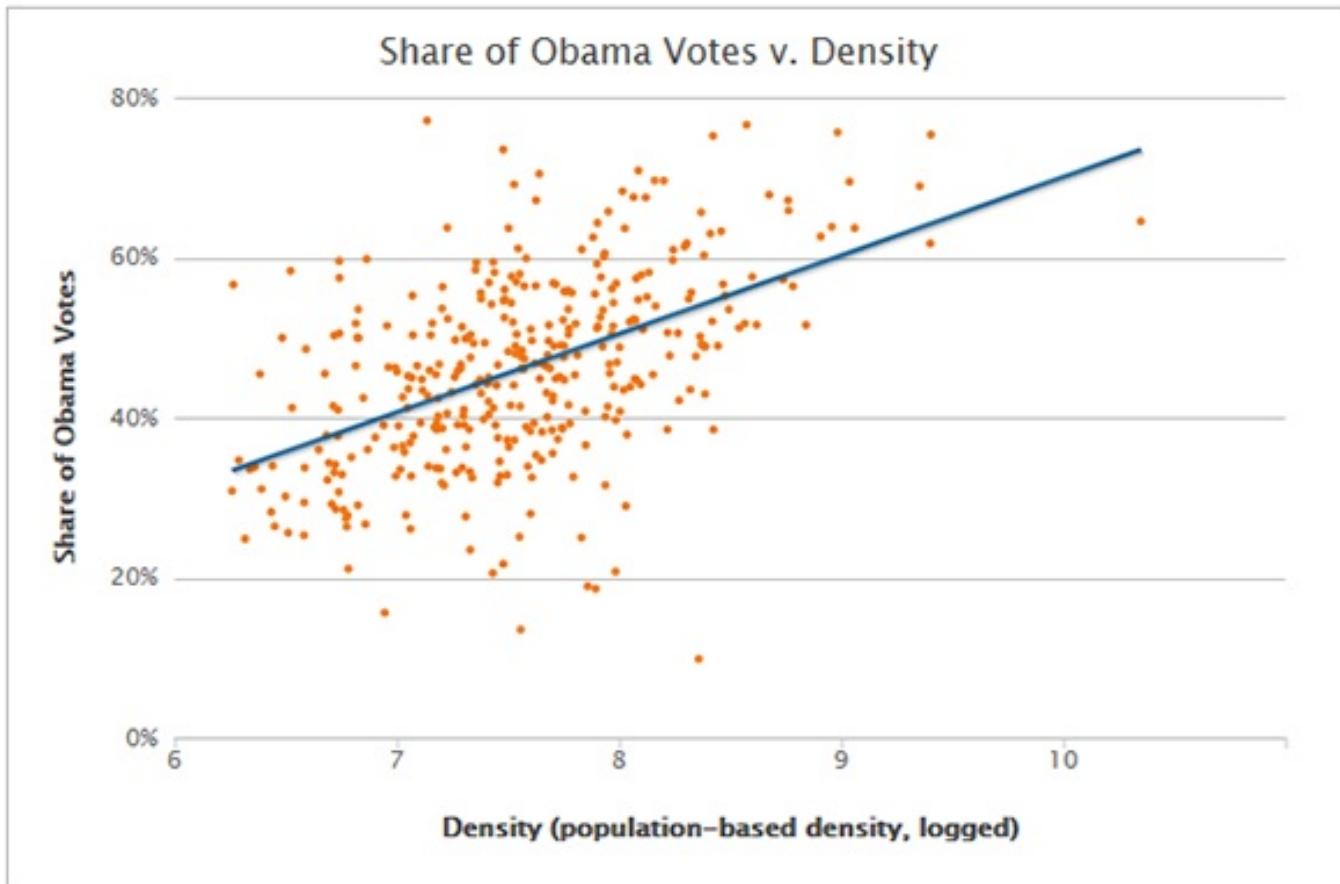
countries furthest to the top-right on the WVS diagram, Sweden, Norway and Denmark all have about 85% urban populations. Go down to the three countries at the bottom left – Jordan, Morocco, and Zimbabwe – and despite Jordan's anomalously high level they're still averaging about 55%.

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](#):



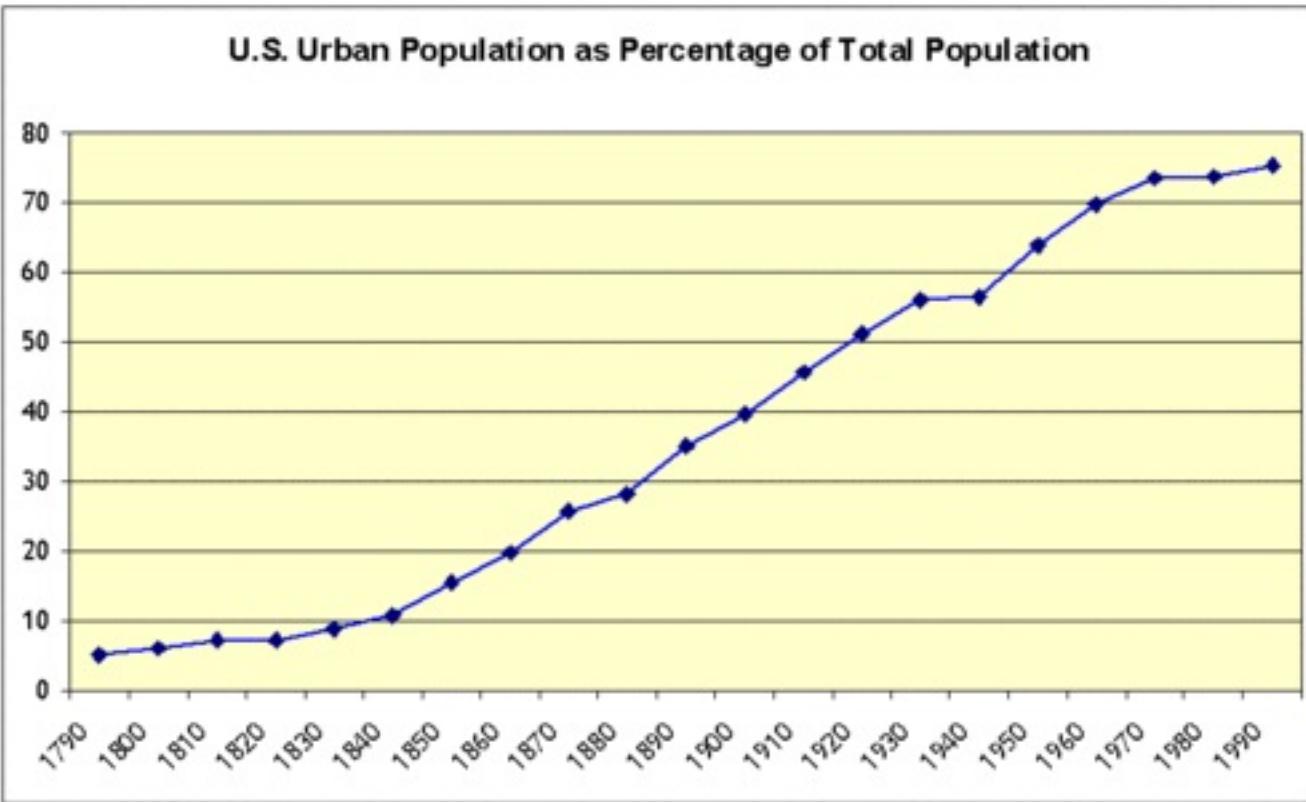
Rural counties – those with <200 people per square mile - lean red at about 65%. Once they pass that 200 person mark, they very quickly start leaning blue until the densest areas - true cities, approach 90% Democratic. Or as Dave Troy notes, "98% of the 50 most dense counties voted Obama. 98% of the 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 [they](#) also 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?

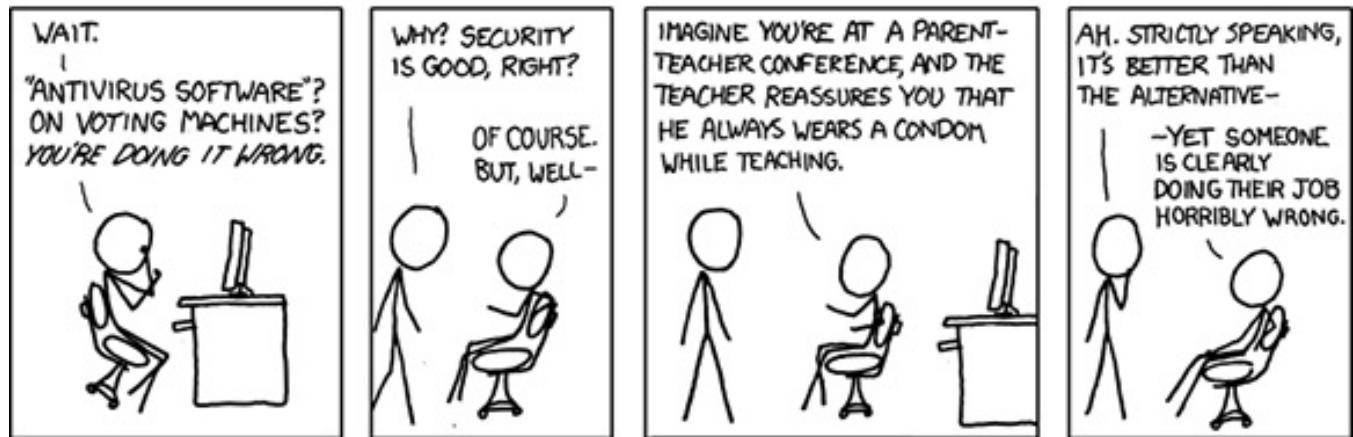
This is one of Moldbug's proposals, and although I think it's been blown out of proportion and he's probably a little embarrassed by it now, it gets brought up enough to be worth addressing.

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...

PREMIER ELECTION SOLUTIONS (FORMERLY DIEBOLD)
HAS BLAMED OHIO VOTING MACHINE ERRORS ON PROBLEMS
WITH THE MACHINES' McAFFEE ANTIVIRUS SOFTWARE.



...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 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 issue (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 gimedats, 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?

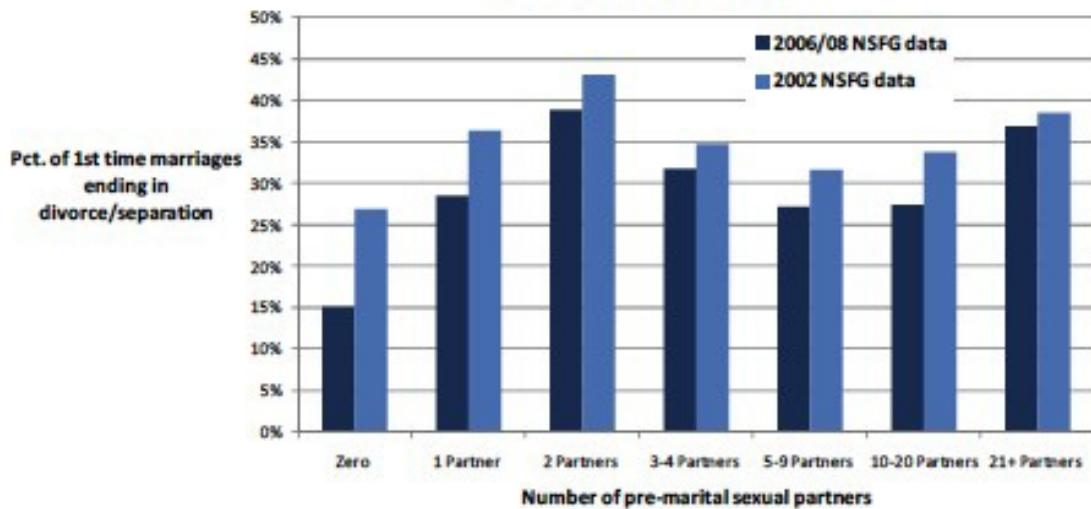
So okay. They have two just-so stories. I can come up with just-so stories too! Like – if a woman sleeps with a lot of people before marriage, she'll be better able to estimate how compatible she is with any given partner. Or – if a woman can sleep with men before marriage, she won't be compelled by 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](#):

Chart 3: Women who have more pre-marital sexual partners are more likely to fail at first marriage



Base: Women aged 15-45 who are/have been married at least once

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.

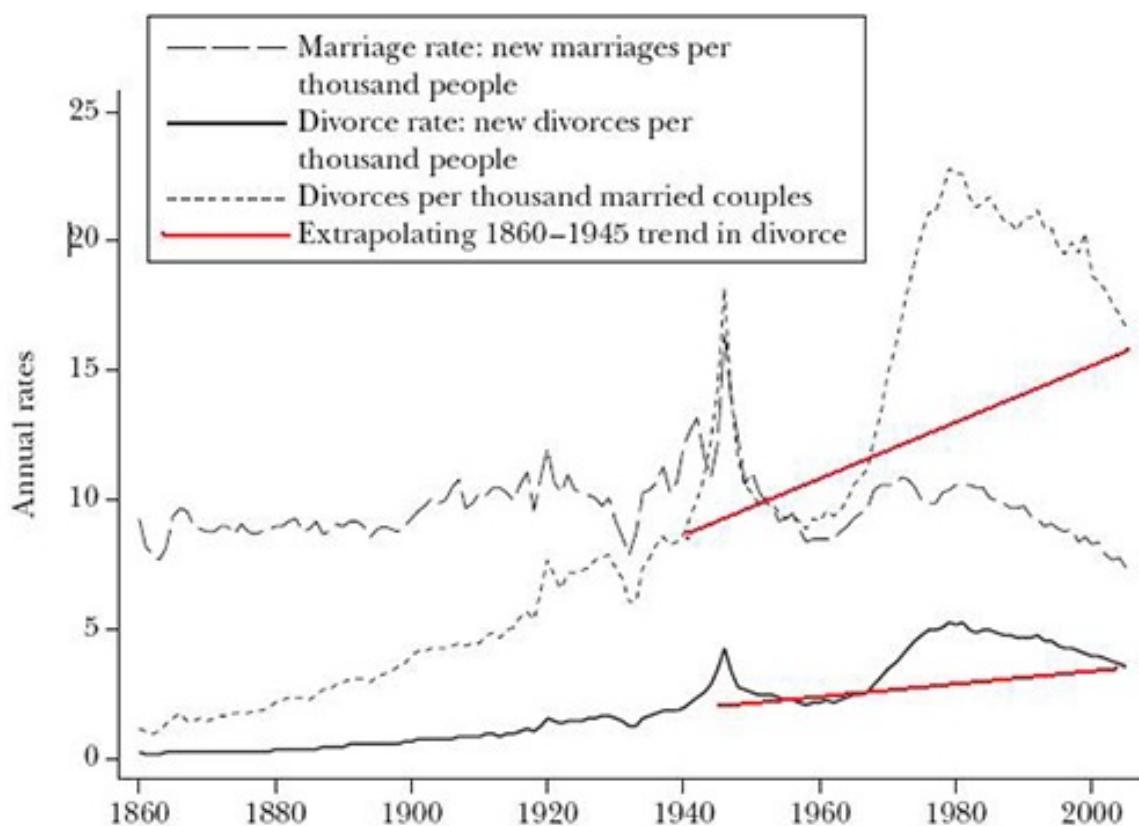
If you phrase the problem as “Those women make stupid and counterproductive sexual choices, how can we shift the balance of power toward men?”, even the sane 30%-or-so of feminists will ignore and oppose you, and with good reason.

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.

Marriages and Divorces per Thousand People, United States 1860–2005



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 more so 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.

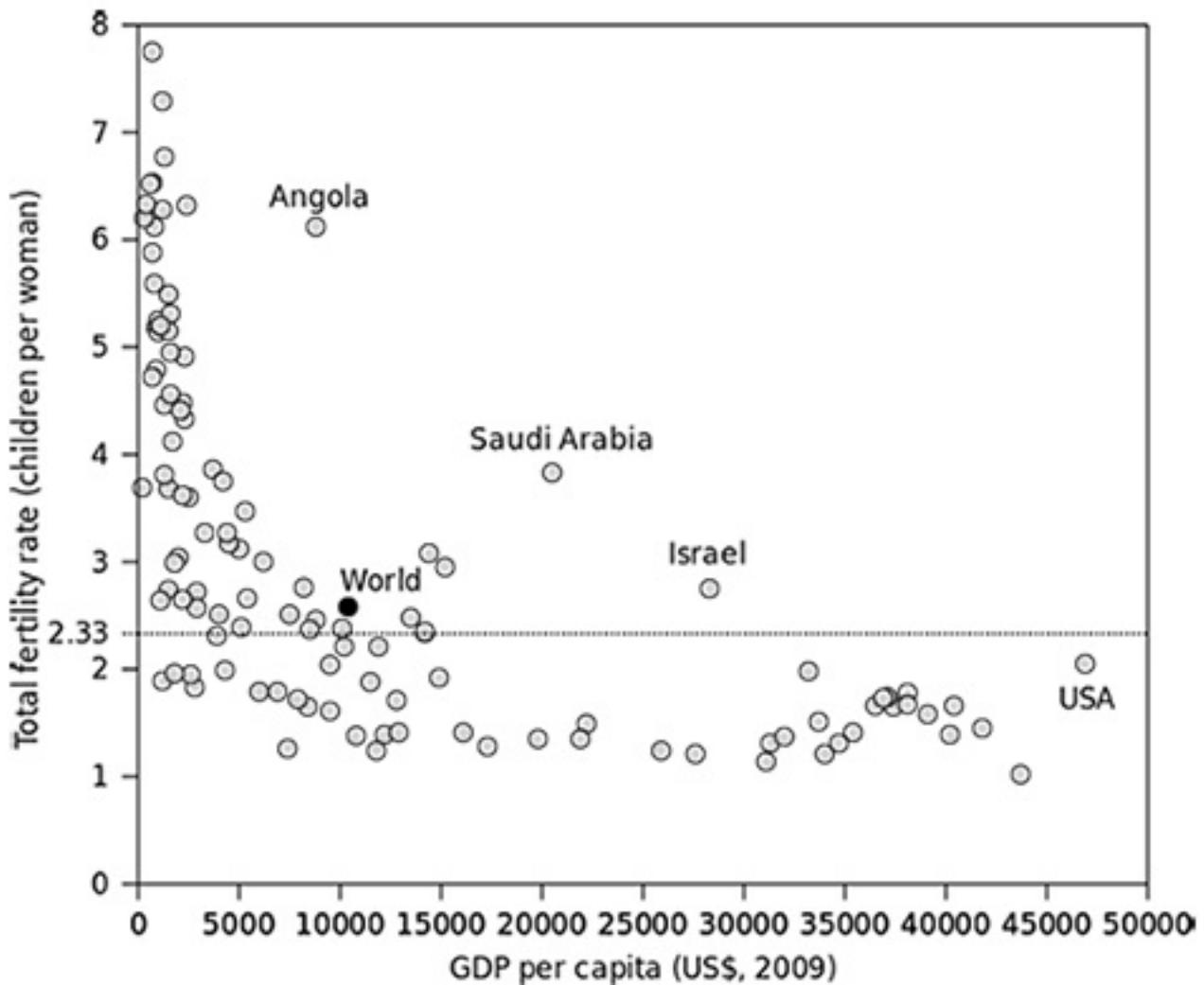
The new norms created by the memetic immune system are exactly the progressive values that Reactionaries blame for the damage: marrying later, trying more partners, using more contraception, having fewer children.

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.

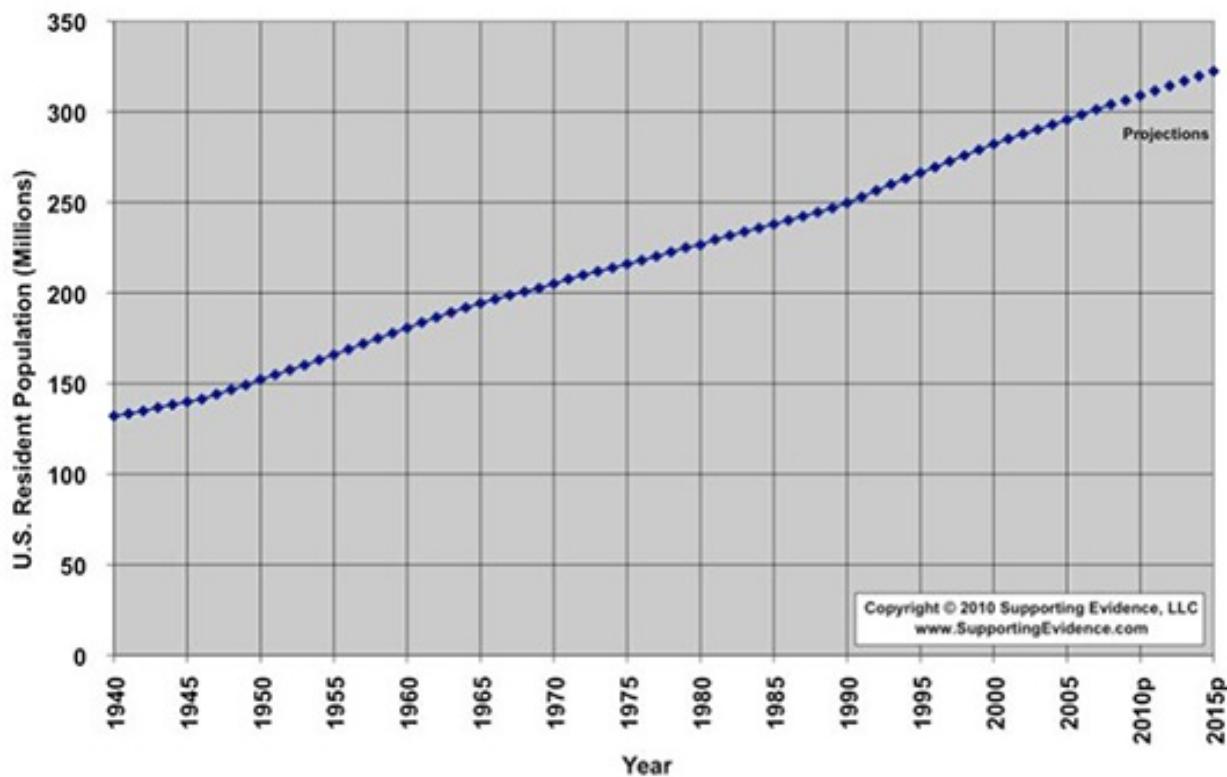
5.3. Are we headed for a demographic catastrophe?

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:

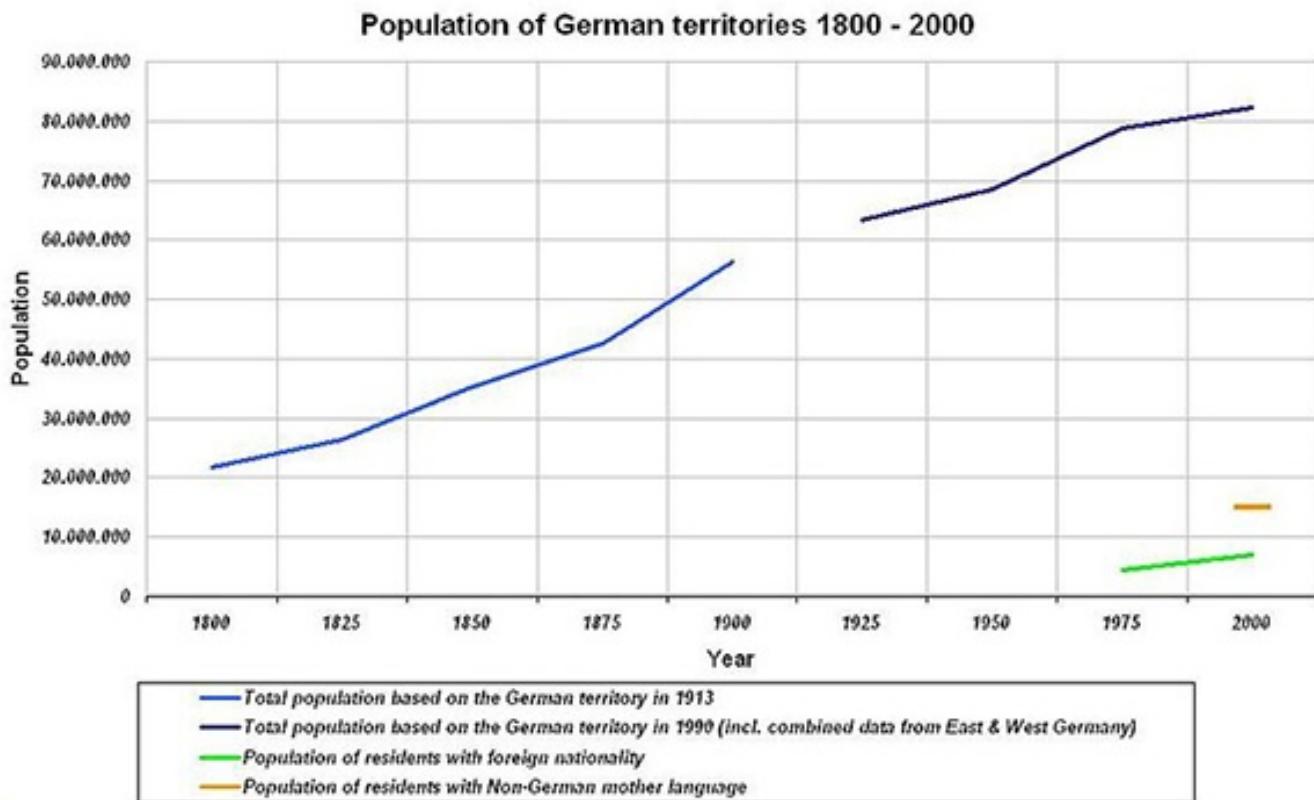


And before we worry about the United States experiencing demographic collapse and tumbleweeds rolling through the streets of New York City, let's double-check to make sure that US population isn't a near-perfectly straight upward-trending line:

U.S. Resident Population Estimates/Projections



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 to 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.

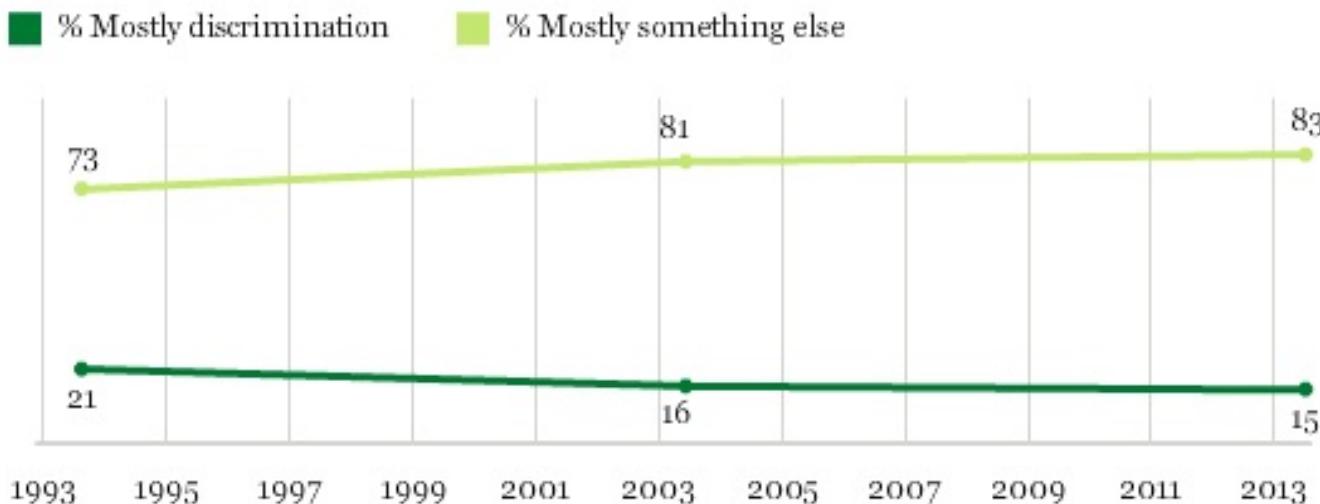
If, as the scientific racists suggest, black people have an average IQ of 85 compared to the white average of 100, then there is *still* a pretty big civil rights battle to be fought getting the average black person to do as well as the average white person with IQ 85. After controlling for IQ, the average 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](#):

Non-Hispanic Whites' Views of Discrimination

On the average, blacks have worse jobs, income, and housing than whites. Do you think this is mostly due to discrimination against blacks, or is it mostly due to something else?



June 13-July 5, 2013

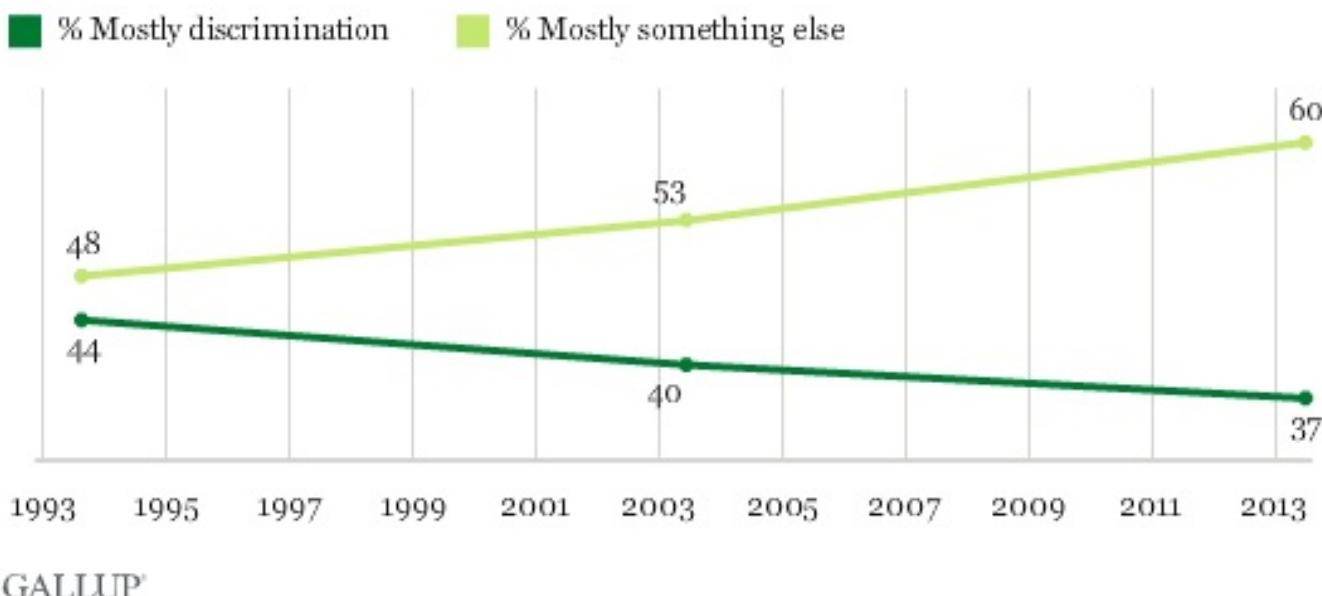
GALLUP®

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?

Blacks' Views of Discrimination

On the average, blacks have worse jobs, income, and housing than whites. Do you think this is mostly due to discrimination against blacks, or is it mostly due to something else?



GALLUP

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), (~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 causal 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 want 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.

Moloch



“*There is no machinery extant, or even rigorously imaginable, that can sustain a single iota of attained value outside the forges of Hell.*

”

— Nick Land

☆☆

[\[REPOST\] The Non-Libertarian FAQ](#)
[Growing Children For Bostrom's Disneyland](#)

[Book Review: Age of Em](#)

[Meditations On Moloch](#)

[Book Review: Red Plenty](#)

[The Goddess of Everything Else](#)

[REPOST] The Non-Libertarian FAQ

Posted on February 22, 2017 by Scott Alexander



This is a repost of the Non-Libertarian FAQ (aka "Why I Hate Your Freedom"), which I wrote about five years ago and which used to be hosted on my website. It no longer completely reflects my current views. I don't think I've switched to believing anything on here is outright false, but I've moved on to different ways of thinking about certain areas. I'm reposting it by popular request and for historical interest only. I've made some very small updates, mostly listing rebuttals that came out over the past few years. I haven't updated the statistics and everything is accurate as of several years ago. I seem to have lost the sources of my images, and I'm sorry; if I've used an image of yours, please let me know and I'll cite you.

Introduction

0.1: 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.

0.2: 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.

0.3: 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.

0.4: 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](#).

0.5: 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 benefited 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 benefiting – I get widgets – and they're benefiting – 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 *mutatis 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 constituents. 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 make 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 Deserts 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 0.4 and 0.6. 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 0.4 and 0.6) was 0.47. But other countries in the study had income elasticity as low as 0.15 (Denmark), 0.16 (Australia), 0.17 (Norway), and 0.19 (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 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 0.4 and 0.6 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 countries).

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 allotted a different number of non-transferable 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, there'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.

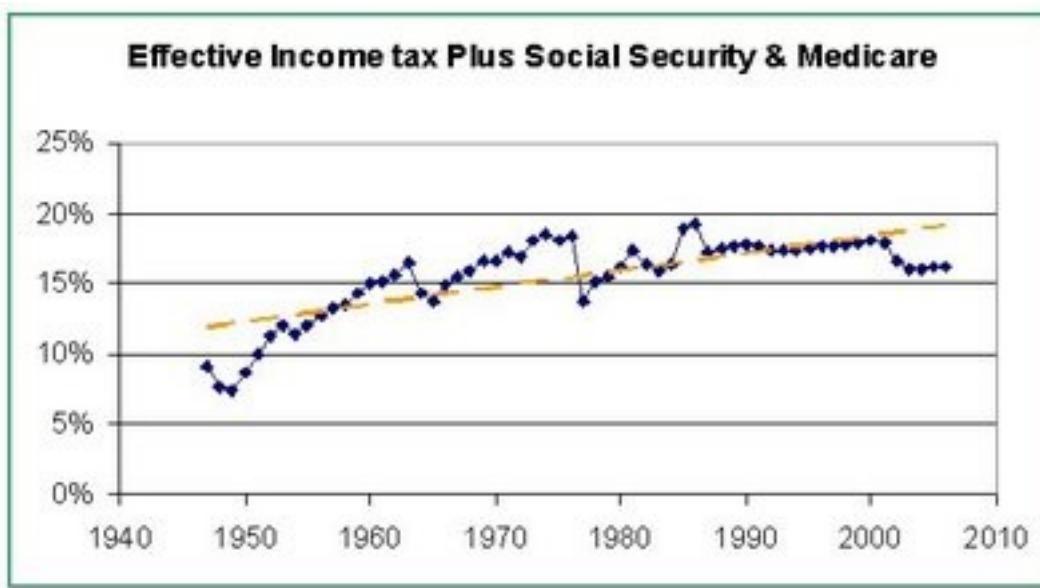
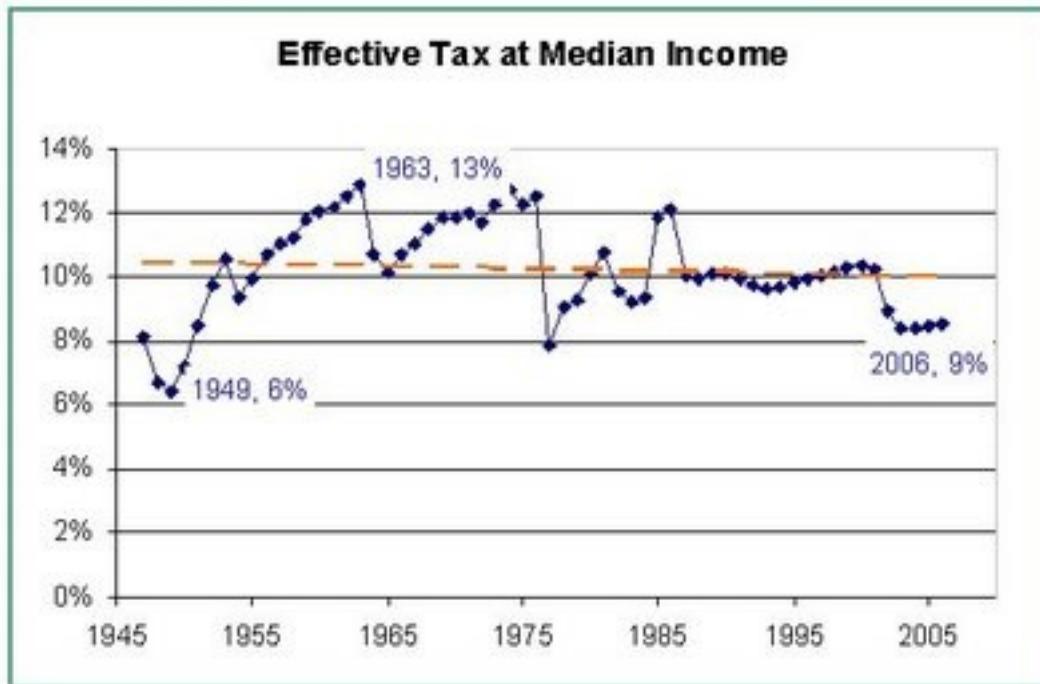
6.3: Taxes are too high.

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.

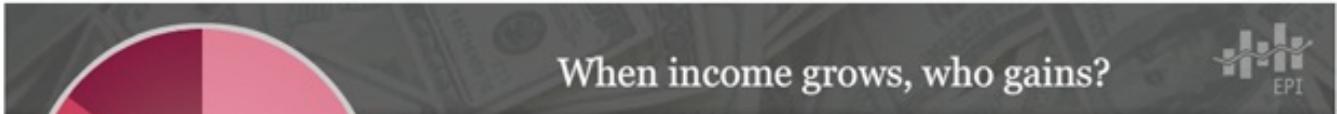
The [Laffer curve](#) certainly exists, but the consensus is that we're still well on the left half of it.

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.



Between 1969 and 2008:



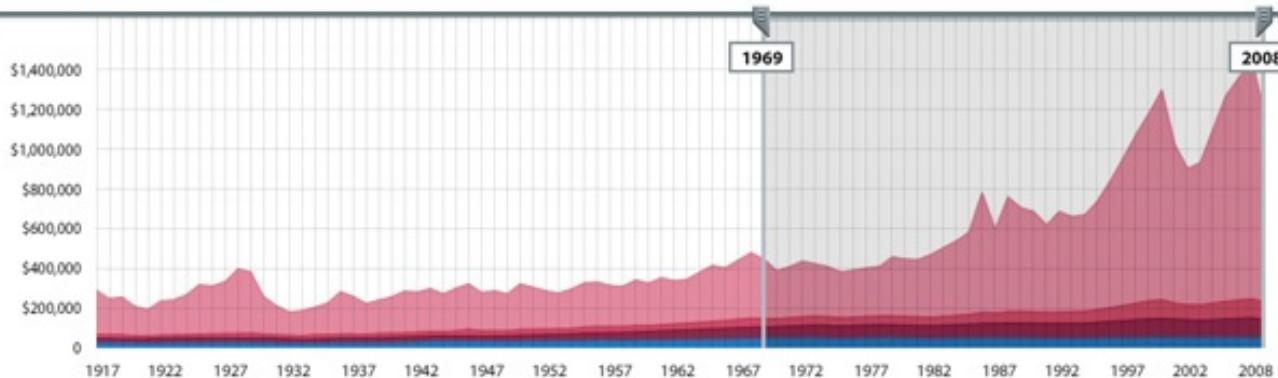
Average incomes in the U.S. grew by \$11,684

All growth went to the richest 10%.

Income for the bottom 90% declined.

■ Bottom 90% ■ Top 5-10% ■ Top 1-5% ■ Top 1% ■ Indicates Decline

Average income over time



Source: The data come from this table: <http://www.econ.berkeley.edu/~saez/TabFig2008.xls> on Emmanuel Saez's website at University of California, Berkeley. **Methodology**

"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 0.000003% of the problem.

So "damnit, 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.

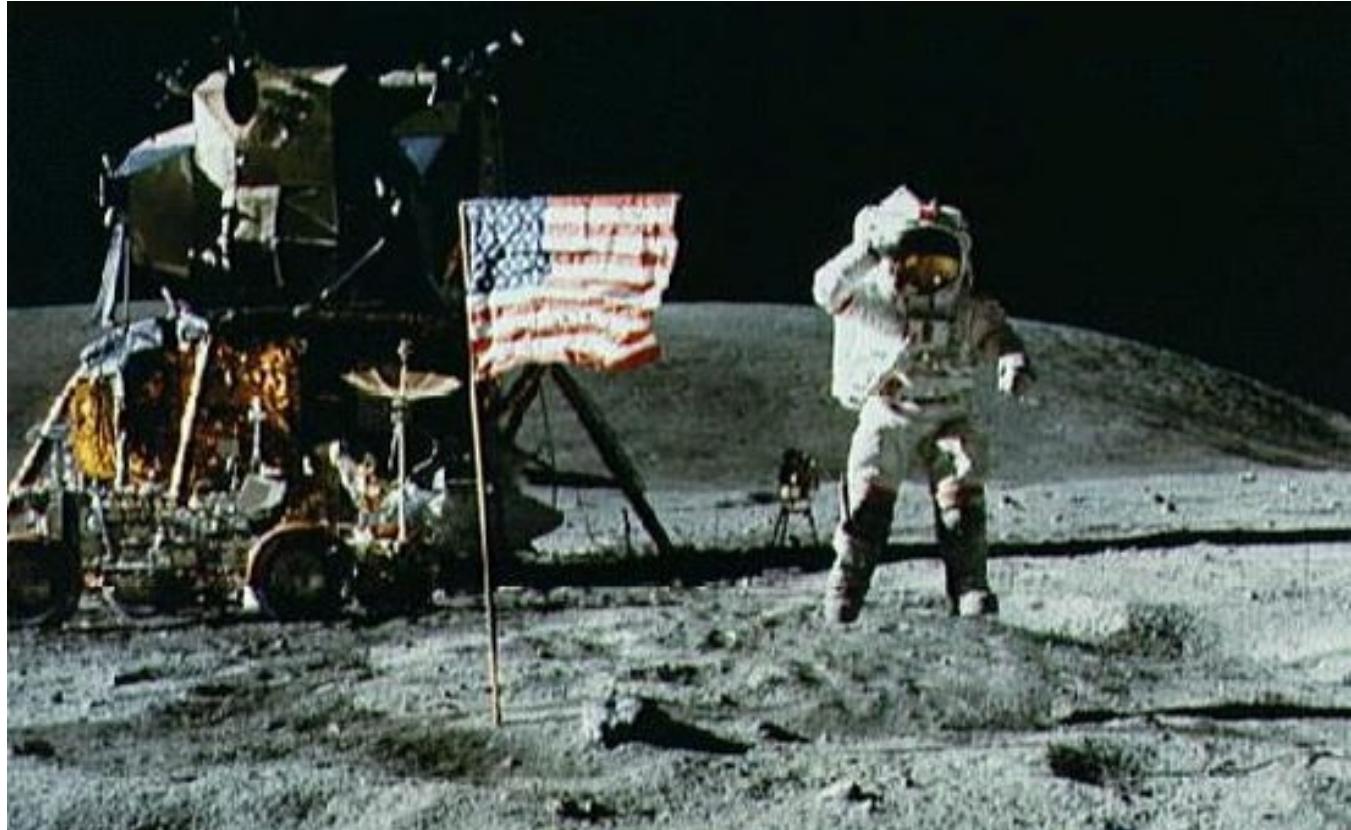
The Counterargument: *Government sometimes, though by no means always, does things right, and some of its institutions and programs are justifiably considered models of efficiency and human ingenuity. There are various reasons why people are less likely to notice these.*

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 is 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, volcanoes, 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 these 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, and 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.

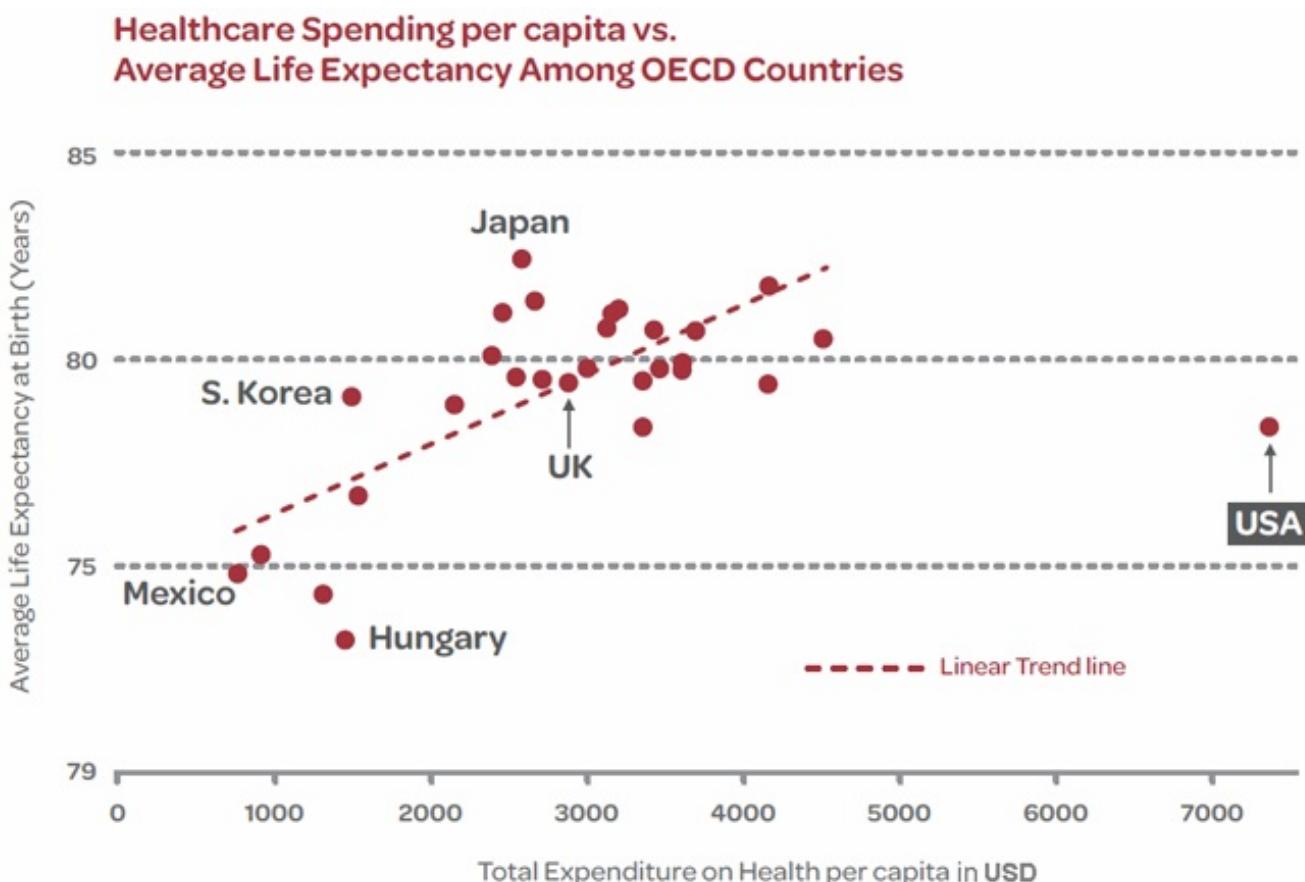
8.2. Government-run health care would be bloated, bureaucratic, and unnecessarily expensive, as opposed to the sleek, efficient service we get from the free market.

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 children's 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.

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 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 seem 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 Alzheimer's 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 Alzheimer's 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 Alzheimer's 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 Misérables*, 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 it 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 difficult for them and bribes biologists not to work with them. They're 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 a 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?

David Friedman wrote a short response [here](#).

Bryan Caplan wrote a response to some of the points about labor [here](#).

Sarah wrote a longer rebuttal here: [Why You Shouldn’t Hate My Freedom](#).

And Nintil wrote another long rebuttal here: [The Non-Non Libertarian FAQ](#).

If you’ve written another rebuttal or you know of one, email me 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](#).

Growing Children For Bostrom's Disneyland

Posted on July 13, 2014 by Scott Alexander



Epistemic status: Started off with something to say, gradually digressed, fell into total crackpottery. Everything after the halfway mark should have been written as a science fiction story instead, but I'm too lazy to change it.

*
*
*

I'm working my way through Nick Bostrom's [Superintelligence: Paths, Dangers, Strategies](#). Review possibly to follow. But today I wanted to write about something that jumped out at me. Page 173. Bostrom is talking about a "multipolar" future similar to Robin Hanson's "em" scenario. The future is inhabited by billions to trillion of vaguely-human-sized agents, probably digital, who are stuck in brutal Malthusian competition with one another.

Hanson tends to view this future as not necessarily so bad. I tend to think Hanson is crazy. I have told him this, and [we have argued about it](#). In particular, I'm pretty sure that brutal Malthusian competition combined with ability to self-edit and other-edit minds necessarily results in paring away everything not directly maximally economically productive. And a lot of things we like – love, family, art, hobbies – are not directly maximally economic productive. Bostrom hedges a lot – appropriate for his line of work – but I get the feeling that he not only agrees with me, but one-ups me by worrying that consciousness itself may not be directly maximally economically productive. He writes:

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.

I think a large number of possible futures converge here (though certainly not all of them, I myself find singleton scenarios more likely) so it's worth asking how doomed we are when we come to this point. Likely we are pretty doomed, but I want to bring up a very faint glimmer of hope in an unexpected place.

It's important to really get our heads around what it means to be in a maximally productive superintelligent Malthusian economy, so I'm going to make some assertions. Instead of lengthy defenses of each, if you disagree with any in particular you can challenge me about it in the comments.

- Every agent is in direct competition with many other entities for limited resources, and ultimately for survival.
- This competition can occur on extremely short (maybe sub-microsecond) time scales.
- A lot of the productive work (and competition) is being done by nanomachines, or if nanomachines are impossible, the nearest possible equivalent.
- Any agent with a disadvantage in any area (let's say intelligence) not balanced by another advantage has already lost and will be outcompeted.
- Any agent that doesn't always take the path that maximizes its utility (defined in objective economic terms) will be outcompeted by another that does.

- Utility calculations will likely be made not according to the vague fuzzy feelings that humans use, but very explicitly, such that agents will know what path maximizes their utility at any given time and their only choice will be to do that or to expect to be outcompeted.
- Agents can only survive a less than maximally utility-maximizing path if they have some starting advantage that gives them a buffer. But gradually these pre-existing advantages will be used up, or copied by the agent's descendants, or copied by other agents that steal them. Things will regress to the pre-existing Malthusianism.

Everyone will behave [perfectly optimally](#), which of course [is terrible](#). It would mean either the total rejection of even the illusion of free will, or free will turning into a simple formality ("You can pick any of these choices you want, but unless you pick Choice C you die instantly.")

The actions of agents become dictated by the laws of economics. Goodness only knows what sort of supergoals these entities might have – maximizing their share of some currency, perhaps a universal currency based on mass-energy? In the first million years, some agent occasionally chooses to violate the laws of economics, and collect less of this currency than it possibly could have because of some principle, but these agents are quickly selected against and go extinct. After that, it's total and invariable. Eventually the thing bumps up against fundamental physical limits, there's no more technological progress to be had, and although there may be some cyclic changes teleological advancement stops.

For me the most graphic version of this scenario is one where all of the interacting agents are very small, very very fast, and with few exceptions operate entirely on reflex. It might look like some of the sci-fi horror ideas of "grey goo". When I imagine things like *that*, the distinction between economics and harder sciences like physics or chemistry starts to blur.

If somehow we captured a one meter sphere of this economic soup, brought it to Earth inside an invincible containment field, and tried to study it, we would probably come up with some very basic laws that it seemed to follow, based on the aggregation of all the entities within it. It would be very silly to try to model the exact calculations of each entity within it – assuming we could even see them or realize they are entities at all. It would just be a really weird volume of space that seemed to follow different rules than our own.

Sci-fi author Karl Schroeder had a term for the post-singularity parts of some of his books – Artificial Nature. That strikes me as exactly right. A hyperproductive end-stage grey goo would take over a rapidly expanding area of space in which all that hypothetical outsiders might notice (non-hypothetical outsiders, of course, would be turned into goo) would be that things are following weird rules and behaving in novel ways.

There's no reason to think this area of space would be homogenous. Because the pre-goo space likely contained different sorts of terrain – void, asteroids, stars, inhabited worlds – different sorts of economic activity would be most productive in each niche, leading to slightly different varieties of goo. Different varieties of goo might cooperate or compete with each other, there might be population implosions or explosions as new resources are discovered or used up – and all of this wouldn't look like economic activity at all to the outside observer. It would look like a weird new kind of physics was in effect, or perhaps like a biological system with different "creatures" in different niches. Occasionally the goo might spin off macroscopic complex objects to fulfill some task those objects could fulfill better than goo, and after a while those objects would dissolve back into the substratum.

Here the goo would fulfill a role a lot like micro-organisms did on Pre-Cambrian Earth – which was also intense Malthusian competition at microscopic levels on short time-scales. Unsurprisingly, the actions of micro-organisms can look physical or chemical to us – put a plate of agar outside and it mysteriously develops white spots. Put a piece of bread outside and it mysteriously develops greenish white spots. Apply the greenish-white spots from the bread to the white spots on the agar, and some of them mysteriously die. Try it too many times and it stops working. It's totally possible to view this on a "guess those are laws of physics" level as well as a "we can dig down and see the terrifying

war-of-all-against-all that emergently results in these large-level phenomena” level.

In this sort of scenario, the only place for consciousness and non-Malthusianism to go would be *higher level structures*.

One of these might be the economy as a whole. Just as ant colonies seem a lot more organism-like than individual ants, so the cosmic economy (or the economies around single stars, if lightspeed limits hold) might seem more organism-like than any of its components. It might be able to sense threats, take actions, or debate very-large-scale policies. If we agree that end-stage-goo is more like biology than like normal-world economics, whatever sort of central planning it comes up with might look more like a brain than like a government. If the components were allowed to plan and control the central planner in detail it would probably be maximally utility maximizing, ie stripped of consciousness and deterministic, but if it arose from a series of least-bad game theoretic bargains it might have some wiggle room.

But I think emergent patterns in the goo itself might be much more interesting.

In the same way our own economy mysteriously pumps out business cycles, end-stage-goo might have cycles of efflorescence and sudden decay. Or the patterns might be weirder. Whorls and eddies in economic activity arising spontaneously out of the interaction of thousands of different complicated behaviors. One day you might suddenly see an extraordinarily complicated mandala or snowflake pattern, like the kind you can get certain variants of Conway’s Game Of Life to make, arise and dissipate.



Source: Latent in the structure of mathematics

Or you might see a replicator. Another thing you can convince Conway’s Game of Life to make.

If the deterministic, law-abiding, microscopically small, instantaneously fast rules of end-stage-goo can be thought of as *pretty much* just a new kind of physics, maybe this kind of physics will allow replicating structures in the same way that normal physics does.

None of the particular economic agents would feel like they were contributing to a replicating pattern, any more than I feel like I’m contributing to [a power law of blogs](#) every time I update here. And it wouldn’t be a disruption in the imperative to only perform the most economically productive action – it would be a pattern that supervenes on everyone’s economically productive behavior.

But it would be creating replicators. Which would eventually retread important advances like sex and mutation and

survival of the fittest and multicellularity and eventually, maybe, sapience.

We would get a whole new meaning of *homo economicus* – but also pan economicus, and mus economicus, and even caenorhabditis economicus.

I wonder what life would be like for those entities. Probably a lot like our own lives. They might be able to manipulate the goo the same way we manipulate normal matter. They might have science to study the goo. They might eventually figure out its true nature, or they might go their entire lifespan as a species without figuring out anything beyond that it has properties it likes to follow. Maybe they would think those properties are the hard-coded law of the universe.

(Here I should pause to point out that none of this requires literal goo. Maybe there is an economy of huge floating asteroid-based factories and cargo-freighters, with Matrioshka brains sitting on artificial planets directing them. Doesn't matter. The patterns in there are harder to map to normal ways of thinking about physics, but I don't see why they couldn't still produce whorls and eddies and replicators.)

Maybe one day these higher-level-patterns would achieve their own singularity, and maybe it would go equally wrong, and they would end up in a Malthusian trap too, and eventually all of their promise would dissipate into extremely economically productive nanomachines competing against one another.

Or they might get a different kind of singularity. Maybe they end up with a paperclip-maximizing singleton. I would think it much less likely that the same kind of complex patterns would arise in the process of paperclip maximization, but maybe they could.

Or maybe, after some number of levels of iteration, they get a positive singularity, a singleton clears up their messes, and they continue studying the universe as superintelligences. Maybe they figure out pretty fast exactly how many levels of entities are beneath them, how many times this has happened before.

I'm not sure if it would be physically possible for them to intervene on the levels below them. In theory, everything beneath them ought to already be literally end-stage. But it might also be locked in some kind of game-theoretic competition that made it less than maximally productive. And so the higher-level entities might be able to design some kind of new matter that outcompetes it and is subject to their own will.

(unless the lower-level systems retained enough intelligence to figure out what was going on, and enough coordinatedness to stop it)

But why would they want to? To them, the lower levels are just physics; always have been, always will be. It would be like a human scientist trying to free electrons from the tyrannous drudgery of orbiting nuclei. Maybe they would sit back and enjoy their victory, sitting at the top of a pyramid of unknown dozens or hundreds of levels of reality.

(Also, just once I want to be able to do armchair futurology without wondering how many times something has already happened.)

Book Review: Age of Em

Posted on May 28, 2016 by Scott Alexander



Note: I really liked this book and if I criticize it that's not meant as an attack but just as what I do with interesting ideas. Note that Robin has offered to debate me about some of this and I've said no – mostly because I hate real-time debates and have bad computer hardware – but you may still want to take this into account when considering our relative positions. Mild **content warning** for murder, rape, and existential horror. Errors in Part III are probably my own, not the book's.

I

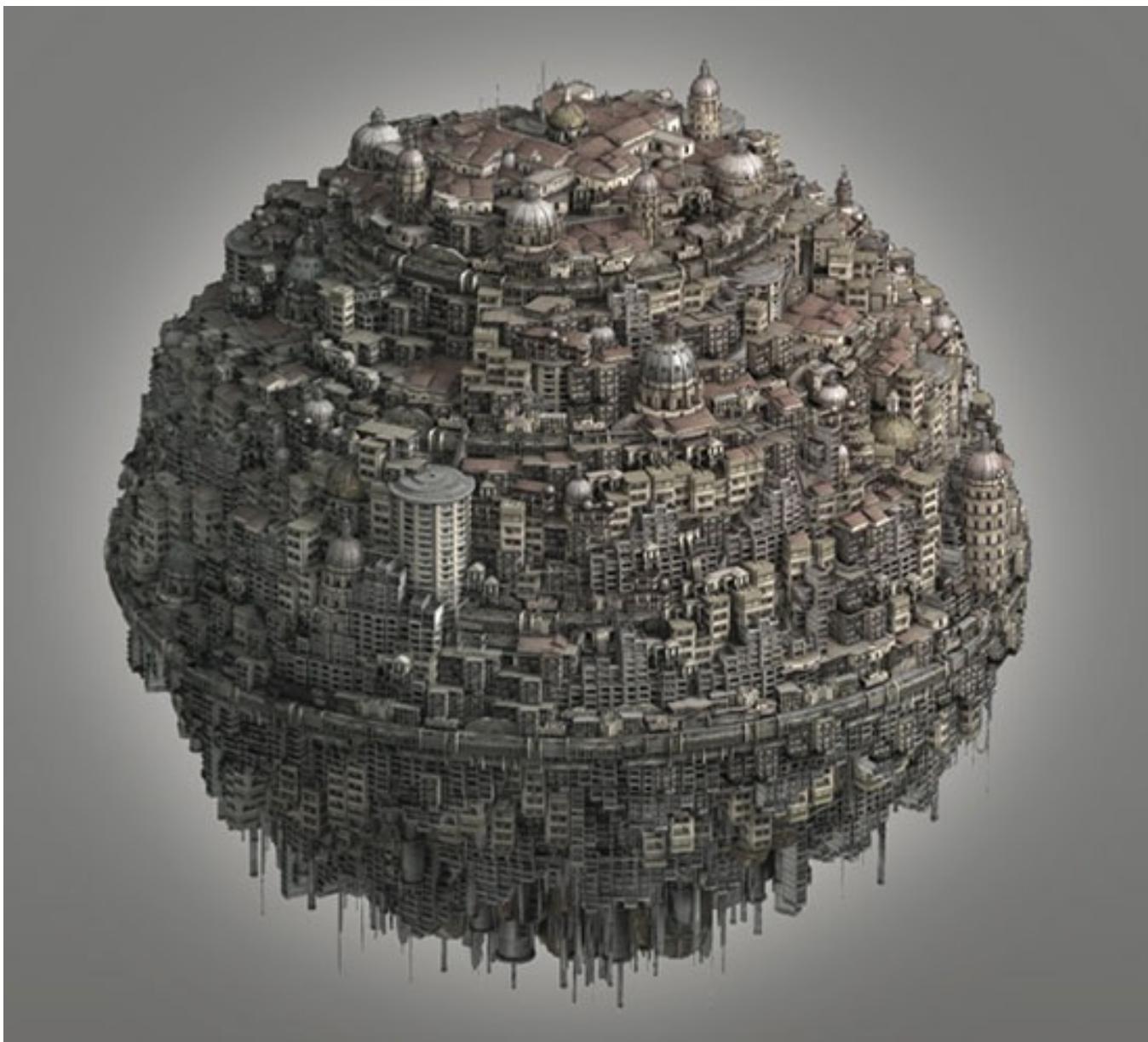
There are some people who are destined to become adjectives. Pick up a David Hume book you've never read before and it's easy to recognize the ideas and style as Humean. Everything Tolkien wrote is Tolkiensque in a non-tautological sense. This isn't meant to denounce either writer as boring. Quite the opposite. They produced a range of brilliant and diverse ideas. But there was a hard-to-define and very consistent ethos at the foundation of both. Both authors were *very much like themselves*.

Robin Hanson is more like himself than anybody else I know. He's obviously brilliant – a PhD in economics, a masters in physics, work for DARPA, Lockheed, NASA, George Mason, and the Future of Humanity Institute. But his greatest aptitude is in being really, *really* [Hansonian](#). Bryan Caplan describes it as well as anybody:

When the typical economist tells me about his latest research, my standard reaction is 'Eh, maybe.' Then I forget about it. When Robin Hanson tells me about his latest research, my standard reaction is 'No way! Impossible!' Then I think about it for years.

This is my experience too. I think I said my first "No way! Impossible!" sometime around 2008 after reading his blog [Overcoming Bias](#). Since then he's influenced my thinking more than almost anyone else I've ever read. When I heard he was writing a book, I was – well, I couldn't even imagine a book by Robin Hanson. When you read a thousand word blog post by Robin Hanson, you have to sit down and think about it and wait for it to digest and try not to lose too much sleep worrying about it. A whole book would be *something*.

I have now read [Age Of Em](#) ([website](#)) and it is indeed something. Even the cover gives you a weird sense of sublimity mixed with unease:



And in this case, judging a book by its cover is entirely appropriate.

II

Age of Em is a work of futurism – an attempt to predict what life will be like a few generations down the road. This is not a common genre – I can't think of another book of this depth and quality in the same niche. Predicting the future is notoriously hard, and that seems to have so far discouraged potential authors and readers alike.

Hanson is not discouraged. He writes that:

Some say that there is little point in trying to foresee the non-immediate future. But in fact there have been many successful forecasts of this sort. For example, we can reliably predict the future cost changes for devices such as batteries or solar cells, as such costs tend to follow a power law of the cumulative device production (Nagy et al 2013). As another example, recently a set of a thousand published technology

forecasts were collected and scored for accuracy, by comparing the forecasted date of a technology milestone with its actual date. Forecasts were significantly more accurate than random, even forecasts 10 to 25 years ahead. This was true separately for forecasts made via many different methods. On average, these milestones tended to be passed a few years before their forecasted date, and sometimes forecasters were unaware that they had already passed (Charbonneau et al, 2013).

A particularly accurate book in predicting the future was *The Year 2000*, a 1967 book by Herman Kahn and Anthony Wiener. It accurately predicted population, was 80% correct for computer and communication technology, and 50% correct for other technology (Albright 2002). On even longer time scales, in 1900 the engineer John Watkins did a good job of forecasting many basic features of society a century later (Watkins 1900) [...]

Some say no one could have anticipated the recent big changes associated with the arrival and consequences of the World Wide Web. Yet participants in the Xanadu hypertext project in which I was involved from 1984 to 1993 correctly anticipated many key aspects of the Web [...] Such examples show that one can use basic theory to anticipate key elements of distant future environments, both physical and social, but also that forecasters do not tend to be much rewarded for such efforts, either culturally or materially. This helps to explain why there are relatively few serious forecasting efforts. But make no mistake, it is possible to forecast the future.

I think Hanson is overstating his case. All except Watkins were predicting only 10 – 30 years in the future, and most of their predictions were simple numerical estimates, eg “the population will be one billion” rather than complex pictures of society. The only project here even remotely comparable in scope to Hanson’s is [John Watkins’ 1900 article](#).

Watkins is classically given some credit for broadly correct ideas like “Cameras that can send pictures across the world instantly” and “telephones that can call anywhere in the world”, but of his 28 predictions, I judge only eight as even somewhat correct. For example, I grant him a prediction that “the average American will be two inches taller because of good medical care” even though he then goes on to say in the same sentence that the average life expectancy will be fifty and suburbanization will be so total that building city blocks will be illegal (sorry, John, only in San Francisco). Most of the predictions seem simply and completely false. Watkins believes all animals and insects will have been eradicated. He believes there will be “peas as large as beets” and “strawberries as large as apples” (these are two separate predictions; he is weirdly obsessed with fruit and vegetable size). We will travel to England via giant combination submarine-hovercrafts that will complete the trip in a lightning-fast two days. There will be no surface-level transportation in cities as all cars and walkways have moved underground. The letters C, X, and Q will be removed from the language. Pneumatic tubes will deliver purchases from stores. “A man or woman unable to walk ten miles at a stretch will be regarded as a weakling.”

Where Watkins is right, he is generally listing a cool technology slightly beyond what was available to his time and predicting we will have it. Nevertheless, he is still mostly wrong. Yet this is Hanson’s example of accurate futurology. And he is *right* to make it his example of accurate futurology, because everything else is even worse.

Hanson has no illusions of certainty. He starts by saying that “conditional on my key assumptions, I expect at least 30% of future situations to be usefully informed by my analysis. Unconditionally, I expect at least 10%.” So he is not explicitly overconfident. But in an implicit sense, it’s just *weird* to see the level of detail he tries to predict – for example, he has two pages about what sort of swear words the far future might use. And the book’s style serves to reinforce its weirdness. The whole thing is written in a sort of professorial monotone that changes little from loving descriptions of the sorts of pipes that will cool future buildings (one of Hanson’s [pet topics](#)) to speculation on our descendants’ romantic relationships (key quote: “The per minute subjective value of an equal relation should not fall much below half of the per-minute value of a relation with the best available open source lover”). And it leans heavily

on a favorite Hansonian literary device – the weirdly general statement about something that sounds like it can't possibly be measurable, followed by a curt reference which if followed up absolutely confirms said statement, followed by relentlessly ringing every corollary of it:

Today, mental fatigue reduces mental performance by about 0.1% per minute. As by resting we can recover at a rate of 1% per minute, we need roughly one-tenth of our workday to be break time, with the duration between breaks being not much more than an hour or two (Trougakos and Hideg 2009; Alvanchi et al 2012)... Thus many em tasks will be designed to take about an hour, and many spurs are likely to last for about this duration.

Or:

Today, painters, novelists, and directors who are experimental artists tend to do their best work at roughly ages 46-52, 38-50, and 45-63 respectively, but those ages are 24-34, 29-40, and 27-43, respectively for conceptual artists (Galenson 2006)... At any one time, the vast majority of actual working ems [should be] near a peak productivity subjective age.

Or:

Wars today, like cities, are distributed evenly across all possible war sizes (Cederman 2003).

At some point I started to wonder whether Hanson was putting me on. Everything is just played *too straight*. Hanson even addresses this:

To resist the temptation to construe the future too abstractly, I'll try to imagine a future full of complex detail. One indication that I've been successful in all these efforts will be if my scenario description sounds less like it came from a typical comic book or science fiction movie, and more like it came from a typical history text or business casebook.

Well, count that project a success. The effect is strange to behold, and I'm not sure it will usher in a new era of futurology. But *Age of Em* is great not just as futurology, but as a bunch of different ideas and purposes all bound up in a futurological package. For example: – An introduction to some of the concepts that recur again and again across Robin's thought – for example, [near vs. far mode](#), the [farmer/forager dichotomy](#), [the inside and outside views](#), [signaling](#). Most of us learned these through years reading Hanson's blog *Overcoming Bias*, getting each chunk in turn, spending days or months thinking over each piece. Getting it all out of a book you can read in a couple of days sounds really hard – but by applying them to dozens of different subproblems involved in future predictions, Hanson makes the reader more comfortable with them, and I expect a lot of people will come out of the book with an intuitive understanding of how they can be applied. – A whirlwind tour through almost every science and a pretty good way to learn about the *present*. If you didn't already know that wars are distributed evenly across all possible war sizes, well, read *Age of Em* and you will know that and many similar things besides. – A manifesto. Hanson often makes predictions by assuming that since the future will be more competitive, future people are likely to converge toward optimal institutions. This is a dangerous assumption for futurology – it's the same line of thinking that led Watkins to assume English would abandon C, X, and Q as inefficient – but it's a *great* assumption if you want a chance to explain your ideas of optimal institutions to thousands of people who think they're reading fun science-fiction. Thus, Robin spends several pages talking about how ems may use prediction markets – an information aggregation technique he invented – to make their decisions. In the real world, Hanson has been trying to push these for decades, with [varying levels](#) of success. Here, in the guise of a future society, he can expose a whole new group of people to their advantages – as well as the advantages of something called “combinatorial auctions” which I am still not smart enough to

understand. – A mind-expanding drug. One of the great risks of futurology is to fail to realize how *different* societies and institutions can be – the same way uncreative costume designers make their aliens look like humans with green skin. A lot of our thoughts about the future involve assumptions we've never really examined critically, and Hanson dynamites those assumptions. For page after page, he gives strong arguments why our descendants might be poorer, shorter-lived, less likely to travel long distances or into space, less progressive and open-minded. He predicts little noticeable technological change, millimeter-high beings living in cities the size of bottles, careers lasting fractions of seconds, humans being incomprehensibly wealthy patrons to their own robot overlords. And *all of it makes sense*.

When I read Stross' *Accelerando*, one of the parts that stuck with me the longest were the Vile Offspring, weird posthuman entities that operated a mostly-incomprehensible Economy 2.0 that humans just sort of hung out on the edges of, goggle-eyed. It was a weird vision – but, for Stross, mostly a black box. *Age of Em* opens the box and shows you every part of what our weird incomprehensible posthuman descendants will be doing in loving detail. Even what kind of swear words they'll use.

III

So, what is the Age of Em?

According to Hanson, AI is really hard and won't be invented in time to shape the posthuman future. But sometime a century or so from now, scanning technology, neuroscience, and computer hardware will advance enough to allow emulated humans, or "ems". Take somebody's brain, scan it on a microscopic level, and use this information to simulate it neuron-by-neuron on a computer. A good enough simulation will map inputs to outputs in exactly the same way as the brain itself, effectively uploading the person to a computer. Uploaded humans will be much the same as biological humans. Given suitable sense-organs, effectuators, virtual avatars, or even robot bodies, they can think, talk, work, play, love, and build in much the same way as their "parent". But ems have three very important differences from biological humans.

First, they have no natural body. They will never need food or water; they will never get sick or die. They can live entirely in virtual worlds in which any luxuries they want – luxurious penthouses, glutinous feasts, Ferraris – can be conjured out of nothing. They will have some limited ability to transcend space, talking to other ems' virtual presences in much the same way two people in different countries can talk on the Internet.

Second, they can run at different speeds. While a normal human brain is stuck running at the speed that physics allow, a computer simulating a brain can simulate it faster or slower depending on preference and hardware availability. With enough parallel hardware, an em could experience a subjective century in an objective week. Alternatively, if an em wanted to save hardware it could process all its mental operations *v e r y s l o w l y* and experience only a subjective week every objective century.

Third, just like other computer data, ems can be copied, cut, and pasted. One uploaded copy of Robin Hanson, plus enough free hardware, can become a thousand uploaded copies of Robin Hanson, each living in their own virtual world and doing different things. The copies could even converse with each other, check each other's work, duel to the death, or – yes – have sex with each other. And if having a thousand Robin Hansons proves too much, a quick ctrl-x and you can delete any redundant ems to free up hard disk space for Civilization 6 ([coming out this October!](#))

Would this count as murder? Hanson predicts that ems will have unusually blasé attitudes toward copy-deletion. If there are a thousand other copies of me in the world, then going to sleep and not waking up just feels like delegating back to a different version of me. If you're still not convinced, Hanson's essay [Is Forgotten Party Death?](#) is a typically disquieting analysis of this proposition. But whether it's true or not is almost irrelevant – at least *some* ems will think this way, and they will be the ones who tend to volunteer to be copied for short term tasks that require termination of

the copy afterwards. If you personally aren't interested in participating, the economy [will leave you behind](#).

The ability to copy ems as many times as needed fundamentally changes the economy and the idea of economic growth. Imagine Google has a thousand positions for Ruby programmers. Instead of finding a thousand workers, they can find one very smart and very hard-working person and copy her a thousand times. With unlimited available labor supply, wages plummet to subsistence levels. "Subsistence levels" for ems are the bare minimum it takes to rent enough hardware from Amazon Cloud to run an em. The overwhelming majority of ems will exist at such subsistence levels. On the one hand, if you've got to exist on a subsistence level, a virtual world where all luxuries can be conjured from thin air is a pretty good place to do it. On the other, such starvation wages might leave ems with little or no leisure time.

Sort of. This gets weird. There's an urban legend about a "test for psychopaths". You tell someone a story about a man who attends his mother's funeral. He met a really pretty girl there and fell in love, but neglected to get her contact details before she disappeared. How might he meet her again? If they answer "kill his father, she'll probably come to that funeral too", they're a psychopath – ordinary people would have a mental block that prevents them from even considering such a drastic solution. And I bring this up because after reading *Age of Em* I feel like Robin Hanson would be able to come up with some super-solution even the psychopaths can't think of, some plan that gets the man a threesome with the girl and her even hotter twin sister at the cost of wiping out an entire continent. Everything about labor relations in *Age of Em* is like this.

For example, suppose you want to hire an em at subsistence wages, but you want them 24 hours a day, 7 days a week. Ems probably need to sleep – that's hard-coded into the brain, and the brain is being simulated at enough fidelity to leave that in. But jobs with tasks that don't last longer than a single day – for example, a surgeon who performs five surgeries a day but has no day-to-day carryover – can get around this restriction by letting an em have one full night of sleep, then copying it. Paste the em at the beginning of the workday. When it starts to get tired, let it finish the surgery it's working on, then delete it and paste the well-rested copy again to do the next surgery. Repeat forever and the em never has to get any more sleep than that one night. You can use the same trick to give an em a "vacation" – just give it *one* of them, then copy-paste that brain-state forever.

Or suppose your ems want frequent vacations, but you want them working every day. Let a "trunk" em vacation every day, then make a thousand copies every morning, work all the copies for twenty-four hours, then delete them. Every copy remembers a life spent in constant vacation, and cheered on by its generally wonderful existence it will give a full day's work. But from the company's perspective, 99.9% of the ems in its employment are working at any given moment.

(another option: work the em at normal subjective speed, then speed it up a thousand times to take its week-long vacation, then have it return to work after only one-one-thousandth of a week has passed in real life)

Given that ems exist at subsistence wages, saving enough for retirement sounds difficult, but this too has weird psychopathic solutions. Thousands of copies of the same em can pool their retirement savings, then have all except a randomly chosen one disappear at the moment of retirement, leaving that one with a nest egg thousands of times what it could have accumulated by its own efforts. Or an em can invest its paltry savings in some kind of low-risk low-return investment and reduce its running speed so much that the return on its investment is enough to pay for its decreased subsistence. For example, if it costs \$100 to rent enough computing power to run an em at normal speed for one year, and you only have \$10 in savings, you can rent 1/1000th of the computer for \$0.10, run at 1/1000th speed, invest your \$10 in a bond that pays 1% per year, and have enough to continue running indefinitely. The only disadvantage is that you'll only experience a subjective week every twenty objective years. Also, since other entities are experiencing a subjective week every second, and some of those entities have nukes, probably there will be some kind of big war, someone will nuke Amazon's data centers, and you'll die after a couple of your subjective minutes. But at least you got to retire!

If ems do find ways to get time off the clock, what will they do with it? Probably they'll have really weird social lives. After all, the existence of em copies is mostly funded by companies, and there's no reason for companies to copy-paste any but the best workers in a given field. So despite the literally trillions of ems likely to make up the world, most will be copies of a few exceptionally brilliant and hard-working individuals with specific marketable talents. Elon Musk might go out one day to the bar with his friend, who is also Elon Musk, and order "the usual". The bartender, who is Elon Musk himself, would know exactly what drink he wants and have it readily available, as the bar caters entirely to people who are Elon Musk. A few minutes later, a few Chesley Sullenbergers might come in after a long day of piloting airplanes. Each Sullenberger would have met hundreds of Musks before and have a good idea about which Musk-Sullenberger conversation topics were most enjoyable, but they might have to adjust for circumstances; maybe the Musks they met before all branched off a most recent common ancestor in 2120, but these are a different branch who were created in 2105 and remember Elon's human experiences but not a lot of the posthuman lives that shaped the 2120 Musks' worldviews. One Sullenberger might tentatively complain that the solar power grid has too many outages these days; a Musk might agree to take the problem up with the Council of Musks, which is totally a thing that exist (Hanson calls these sorts of groups "copy clans" and says they are "a natural candidate unit for finance, reproduction, legal, liability, and political representation").

Romance could be even weirder. Elon Musk #2633590 goes into a bar and meets Taylor Swift #105051, who has a job singing in a nice local nightclub and so is considered prestigious for a Taylor Swift. He looks up a record of what happens when Elon Musks ask Taylor Swifts out and finds they are receptive on 87.35% of occasions. The two start dating and are advised by the Council of Musks and the Council of Swifts on the issues that are known to come up in Musk-Swift relationships and the best solutions that have been found to each. Unfortunately, Musk #2633590 is transferred to a job that requires operating at 10,000x human speed, but Swift #105051's nightclub runs at 100x speed and refuses to subsidize her to run any faster; such a speed difference makes normal interaction impossible. The story has a happy ending; Swift #105051 allows Musk #2633590 to have her source code, and whenever he is feeling lonely he spends a little extra money to instantiate a high-speed copy of her to hang out with.

(needless to say, these examples are not exactly word-for-word taken from the book, but they're heavily based off of Hanson's more abstract descriptions)

The em world is not just very weird, it's also very very big. Hanson notes that labor is a limiting factor in economic growth, yet even today the economy doubles about once every fifteen years. Once you can produce skilled labor through a simple copy-paste operation, especially labor you can run at a thousand times human speed, the economy will go through the roof. He writes that:

To generate an empirical estimate of em economy doubling times, we can look at the timescales it takes for machine shops and factories today to make a mass of machines of a quality, quantity, variety, and value similar to that of machines that they themselves contain. Today that timescale is roughly 1 to 3 months. Also, designs were sketched two to three decades ago for systems that might self-replicate nearly completed in 6 to 12 months... these estimates suggest that today's manufacturing technology is capable of self-replicating on a scale of a few weeks to a few months.

Hanson thinks that with further innovation, such times can be reduced so far that "the economy might double every objective year, month, week, or day." As the economy doubles the labor force – ie the number of ems – may double with it, until only a few years after the first ems the population numbers in the trillions. But if the em population is doubling every day, there had better be some pretty amazing construction efforts going on. The only thing that could possibly work on that scale is prefabricated modular construction of giant superdense cities, probably made mostly out of some sort of proto early-stage computronium (plus cooling pipes). Ems would be reluctant to travel from one such city to another – if they exist at a thousand times human speed, a trip on a hypersonic airliner that could go from New York to Los Angeles in an hour would still take forty subjective days. Who wants to be on an airplane for forty days?

(long-distance trade is also rare, since if the economy doubles fast enough it means that by the time goods reach their destination they could be almost worthless)

The real winners of this ultra-fast-growing economy? Ordinary humans. While humans will be way too slow and stupid to do anything useful, they will tend to have non-subsistence amounts of money saved up from their previous human lives, and also be running at speeds thousands of times slower than most of the economy. When the economy doubles every day, so can your bank account. Ordinary humans will become rarer, less relevant, but fantastically rich – a sort of doddering Neanderthal aristocracy spending sums on a cheeseburger that could support thousands of ems in luxury for entire lifetimes. While there will no doubt be pressure to liquidate humans and take their stuff, Hanson hopes that the spirit of rule of law – the same spirit that protects rich minority groups today – will win out, with rich ems reluctant to support property confiscation lest it extend to them also. Also, em retirees will have incentives a lot like humans – they have saved up money and go really slow – and like AARP members today they may be able to obtain disproportionate political power which will then protect the interests of slow rich people.

But we might not have much time to enjoy our sudden rise in wealth. Hanson predicts that the Age of Em will last for subjective em millennia – ie about one to two actual human years. After all, most of the interesting political and economic activity is going on at em timescales. In the space of a few subjective millennia, either someone will screw up and cause the apocalypse, somebody will invent real superintelligent AI that causes a technological singularity, or some other weird thing will happen taking civilization beyond the point that even Robin dares to try to predict.

IV

Hanson understands that people might not like the idea of a future full of people working very long hours at subsistence wages forever (Zack Davis' [Contract-Drafting Em](#) song is, as usual, relevant). But Hanson himself does not view this future as dystopian. Despite our descendants' by-the-numbers poverty, they will avoid the miseries commonly associated with poverty today. There will be no dirt or cockroaches in their sparkling virtual worlds, nobody will go hungry, petty crime will be all-but-eliminated, and unemployment will be low. Anybody who can score some leisure time will have a dizzying variety of hyperadvanced entertainment available, and as for the people who can't, they'll mostly have been copied from people who really like working hard and don't miss it anyway. As unhappy as we moderns may be contemplating em society, ems themselves will not be unhappy! And as for us:

The analysis in this book suggests that lives in the next great era may be as different from our lives as our lives are from farmers' lives, or farmers' lives are from foragers' lives. Many readers of this book, living industrial era lives and sharing industrial era values, may be disturbed to see a forecast of em era descendants with choices and lifestyles that appear to reject many of the values that they hold dear. Such readers may be tempted to fight to prevent the em future, perhaps preferring a continuation of the industrial era. Such readers may be correct that rejecting the em future holds them true to their core values. But I advise such readers to first try hard to see this new era in some detail from the point of view of its typical residents. See what they enjoy and what fills them with pride, and listen to their criticisms of your era and values.

A short digression: there's a certain strain of thought I find infuriating, which is "My traditionalist ancestors would have disapproved of the changes typical of my era, like racial equality, more open sexuality, and secularism. But I am smarter than them, and so totally okay with how the future will likely have values even more progressive and shocking than my own. Therefore I pre-approve of any value changes that might happen in the future as definitely good and better than our stupid hidebound present."

I once read a science-fiction story that depicted a pretty average sci-fi future – mighty starships, weird aliens,

confederations of planets, post-scarcity economy – with the sole unusual feature that rape was considered totally legal, and opposition to such as bigoted and ignorant as opposition to homosexuality is today. Everybody got really angry at the author and said it was offensive for him to even speculate about that. Well, that's the method by which our cheerful acceptance of any possible future values is maintained: restricting the set of "any possible future values" to "values slightly more progressive than ours" and then angrily shouting down anyone who discusses future values that actually sound bad. But of course the *whole question* of how worried to be about future value drift *only makes sense* in the context of future values that genuinely violate our current values. Approving of all future values except ones that would be offensive to even speculate about is the same faux-open-mindedness as tolerating anything except the outgroup.

Hanson deserves credit for positing a future whose values are likely to upset even the sort of people who say they don't get upset over future value drift. I'm not sure whether or not he deserves credit for not being upset by it. Yes, it's got low-crime, ample food for everybody, and full employment. But so does *Brave New World*. The whole *point* of dystopian fiction is pointing out that we have complicated values beyond material security. Hanson is absolutely right that our traditionalist ancestors would view our own era with as much horror as some of us would view an em era. He's even right that on utilitarian grounds, it's hard to argue with an em era where everyone is really happy working eighteen hours a day for their entire lives because we selected for people who feel that way. But at some point, can we make the Lovecraftian argument of "I know my values are provincial and arbitrary, but they're *my* provincial arbitrary values and I will make any sacrifice of blood or tears necessary to defend them, even unto the gates of Hell?"

This brings us to an even worse scenario.

There are a lot of similarities between Hanson's futurology and (my possibly erroneous interpretation of) the futurology of Nick Land. I see Land as saying, like Hanson, that the future will be one of quickly accelerating economic activity that comes to dominate a bigger and bigger portion of our descendants' lives. But whereas Hanson's framing focuses on the participants in such economic activity, playing up their resemblances with modern humans, Land takes a bigger picture. He talks about the economy itself acquiring a sort of self-awareness or agency, so that the destiny of civilization is consumed by the imperative of economic growth.

Imagine a company that manufactures batteries for electric cars. The inventor of the batteries might be a scientist who really believes in the power of technology to improve the human race. The workers who help build the batteries might just be trying to earn money to support their families. The CEO might be running the business because he wants to buy a really big yacht. And the whole thing is there to eventually, somewhere down the line, let a suburban mom buy a car to take her kid to soccer practice. Like most companies the battery-making company is primarily a profit-making operation, but the profit-making-ness draws on a lot of not-purely-economic actors and their not-purely-economic subgoals.

Now imagine the company fires all its employees and replaces them with robots. It fires the inventor and replaces him with a genetic algorithm that optimizes battery design. It fires the CEO and replaces him with a superintelligent business-running algorithm. All of these are good decisions, from a profitability perspective. We can absolutely imagine a profit-driven shareholder-value-maximizing company doing all these things. But it reduces the company's non-masturbatory participation in an economy that points outside itself, limits it to just a tenuous connection with soccer moms and maybe some shareholders who want yachts of their own.

Now take it further. Imagine there are no human shareholders who want yachts, just banks who lend the company money in order to increase their own value. And imagine there are no soccer moms anymore; the company makes batteries for the trucks that ship raw materials from place to place. Every non-economic goal has been stripped away from the company; it's just an appendage of Global Development.

Now take it even further, and imagine this is what's happened everywhere. There are no humans left; it isn't

economically efficient to continue having humans. Algorithm-run banks lend money to algorithm-run companies that produce goods for other algorithm-run companies and so on ad infinitum. Such a masturbatory economy would have all the signs of economic growth we have today. It could build itself new mines to create raw materials, construct new roads and railways to transport them, build huge factories to manufacture them into robots, then sell the robots to whatever companies need more robot workers. It might even eventually invent space travel to reach new worlds full of raw materials. Maybe it would develop powerful militaries to conquer alien worlds and steal their technological secrets that could increase efficiency. It would be vast, incredibly efficient, and utterly pointless. The real-life incarnation of those strategy games where you mine Resources to build new Weapons to conquer new Territories from which you mine more Resources and so on forever.

But this seems to me the natural end of the economic system. Right now it needs humans only as laborers, investors, and consumers. But robot laborers are potentially more efficient, companies based around algorithmic trading are already pushing out human investors, and most consumers already aren't individuals – they're companies and governments and organizations. At each step you can gain efficiency by eliminating humans, until finally humans aren't involved *anywhere*.

True to form, Land doesn't see this as a dystopia – I think he conflates “maximally efficient economy” with “God”, which is a *hell* of a thing to conflate – but I do. And I think it provides an important new lens with which to look at the Age of Em.

The Age of Em is an economy in the early stages of such a transformation. Instead of being able to replace everything with literal robots, it replaces them with humans who have had some aspects of their humanity stripped away. Biological bodies. The desire and ability to have children normally. Robin doesn't think people will lose all leisure time and non-work-related desires, but he doesn't seem too sure about this and it doesn't seem to bother him much if they do.

I envision a spectrum between the current world of humans and Nick Land's Ascended Economy. Somewhere on the spectrum we have ems who get leisure time. A little further on the spectrum we have ems who don't get leisure time.

But we can go further. Hanson imagines that we can “tweak” em minds. We may not understand the brain enough to create totally new intelligences from the ground up, but by his Age of Em we should understand it well enough to make a few minor hacks, the same way even somebody who doesn't know HTML or CSS can usually figure out how to change the background color of a webpage with enough prodding. Many of these mind tweaks will be the equivalent of psychiatric drugs – some might even be computer simulations of what we observe to happen when we give psychiatric drugs to a biological brain. But these tweaks will necessarily be much stronger and more versatile, since we no longer care about bodily side effects (ems don't have bodies) and we can apply it to only a single small region of the brain and avoid actions anywhere else. You could also very quickly advance brain science – the main limits today are practical (it's really hard to open up somebody's brain and do stuff to it without killing them) and ethical (the government might have some words with you if you tried). An Age of Em would remove both obstacles, and give you the added bonus of being able to make thousands of copies of your test subjects for randomized controlled trials, reloading any from a saved copy if they died. Hanson envisions that:

As the em world is a very competitive world where sex is not needed for reproduction, and as sex can be time and attention-consuming, ems may try to suppress sexuality, via mind tweaks that produce effects analogous to castration. Such effects might be temporary, perhaps with a consciously controllable on-off switch... it is possible that em brain tweaks could be found to greatly reduce natural human desires for sex and related romantic and intimate pair bonding without reducing em productivity. It is also possible that many of the most productive ems would accept such tweaks.

Possible? I can do that *right now* with a high enough dose of Paxil, and I don't even have to upload your brain to a

computer first. Fun stories about Musk #2633590 and Swift #105051 aside, I expect this would happen about ten minutes after the advent of the Age of Em, and we would have taken another step down the path to the Ascended Economy.

There are dozens of other such tweaks I can think of, but let me focus on two.

First, stimulants have a very powerful ability to focus the brain on the task at hand, as anybody who's taken Adderall or modafinil can attest. Their main drawbacks are addictiveness and health concerns, but in a world where such pills can be applied as mental tweaks, where minds have no bodies, and where any mind that gets too screwed up can be reloaded from a backup copy, these are barely concerns at all. Many of the purely mental side effects of stimulants come from their effects in parts of the brain not vital to the stimulant effect. If we can selectively apply Adderall to certain brain centers but not others, then unapply it at will, then from employers' point of view there's no reason not to have all workers dosed with superior year 2100 versions of Adderall at all times. I worry that not only will workers not have any leisure time, but they'll be neurologically incapable of having their minds drift off while on the job. Davis' contract-drafting em who starts wondering about philosophy on the job wouldn't get terminated. He would just have his simulated-Adderall dose increased.

Second, Robin managed to write an entire book about emulated minds without using the word "[wireheading](#)". This is another thing we can do right now, with today's technology – but once it's a line of code and not a costly brain surgery, it should become nigh-universal. Give ems the control switches to their own reward centers and all questions about leisure time become irrelevant. Give bosses the control switches to their employees' reward centers, and the situation changes markedly. Hanson says that there probably won't be too much slavery in the em world, because it will likely have strong rule of law, because slaves aren't as productive as free workers, and there's little advantage to enslaving someone when you could just pay them subsistence wages anyway. But slavery isn't *nearly* as abject and inferior a condition as the one where somebody else has the control switch to your reward center. Combine that with the stimulant use mentioned above, and you can have people who will never have nor want to have any thought about anything other than working on the precise task at which they are supposed to be working at any given time.

This is something I worry about even in the context of normal biological humans. But Hanson already believes em worlds will have few regulations and be able to ignore the moral horror of 99% of the population by copying and using the 1% who are okay with something. Combine this with a situation where brains are easily accessible and tweakable, and this sort of scenario becomes horribly likely.

I see almost no interesting difference between an em world with full use of these tweaks and an Ascended Economy world. Yes, there are things that look vaguely human in outline laboring in the one and not the other, but it's not like there will be different thought processes or different results. I'm not even sure what it would mean for the ems to be conscious in a world like this – they're not doing anything interesting with the consciousness. The best we could say about this is that if the wireheading is used liberally it's a lite version of the world where everything gets converted to [hedonium](#).

V

In a book full of weird ideas, there is only one idea rejected as too weird. And in a book written in a professorial monotone, there's only one point at which Hanson expresses anything like emotion:

Some people foresee a rapid local "intelligence explosion" happening soon after a smart AI system can usefully modify its local architecture (Chalmers 2010; Hanson and Yudkowsky 2013; Yudkowsky 2013; Bostrom 2014)... Honestly to me this local intelligence explosion scenario looks suspiciously like a supervillain comic book plot. A flash of insight by a lone genius lets him create a genius AI. Hidden in its super-

villain research lab lair, this genius villain AI works out unprecedented revolutions in AI design, turns itself into a super-genius, which then invents super-weapons and takes over the world. Bwa ha ha.

For someone who just got done talking about the sex lives of uploaded computers in millimeter-tall robot bodies running at 1000x human speed, Robin is sure quick to use the absurdity heuristic to straw-man intelligence explosion scenarios as “comic book plots”. Take away his weird authorial tic of using the words “genius” and “supervillain”, this scenario reduces to “Some group, perhaps Google, perhaps a university, invent an artificial intelligence smart enough to edit its own source code; exponentially growing intelligence without obvious bound follows shortly thereafter”. Yes, it’s weird to think that there may be a sudden quantum leap in intelligence like this, but no weirder than to think most of civilization will transition from human to em in the space of a year or two. I’m a little bit offended that this is the only idea given this level of dismissive treatment. Since I do have immense respect for Robin, I hope my offense doesn’t color the following thoughts too much.

Hanson’s arguments against AI seem somewhat motivated. He admits that AI researchers generally estimate less than 50 years before we get human-level artificial intelligence, a span shorter than his estimate of a century until we can upload ems. He even admits that no AI researcher thinks ems are a plausible route to AI. But he dismisses this by saying when he asks AI experts informally, they say that in their own field, they have only noticed about 5-10% of the progress they expect would be needed to reach human intelligence over the past twenty years. He then multiplies out to say that it will probably take at least 400 years to reach human-level AI. I have two complaints about this estimate.

First, he is explicitly ignoring published papers surveying hundreds of researchers using validated techniques, in favor of what he describes as “meeting experienced AI experts informally”. But even though he feels comfortable rejecting vast surveys of AI experts as potentially biased, as best I can tell he does not ask a single neuroscientist to estimate the date at which brain scanning and simulation might be available. He just says that “it seems plausible that sufficient progress will be made in roughly a century or so”, citing a few hopeful articles by very enthusiastic futurists who are not neuroscientists or scanning professionals themselves and have not talked to any. This seems to me to be an extreme example of isolated demands for rigor. No matter how many AI scientists think AI is soon, Hanson will cherry-pick the surveying procedures and results that make it look far. But if a few futurists think brain emulation is possible, then no matter what anybody else thinks that’s good enough for him.

Second, one would expect that even if there were only 5-10% progress over the last twenty years, then there would be faster progress in the future, since the future will have a bigger economy, better supporting technology, and more resources invested in AI research. Robin answers this objection by saying that “increases in research funding usually give much less than proportionate increases in research progress” and cites Alston et al 2011. I looked up Alston et al 2011, and it is a paper relating crop productivity to government funding of agriculture research. There was no attempt to relate its findings to any field other than agriculture, nor to any type of funding other than government. But studies show that while public research funding often does have minimal effects, the effect of private research funding is usually much larger. A single sentence citing a study in crop productivity to apply to artificial intelligence while ignoring much more relevant results that contradict it seems like a really weak argument for a statement as potentially surprising as “amount of research does not affect technological progress”.

I realize that Hanson has done a lot more work on this topic and he couldn’t fit all of it in this book. I disagree with his other work too, and I’ve said so elsewhere. For now I just want to say that the arguments in this book seem weak to me.

I also want to mention what seems to me a very Hansonian counterargument to the ems-come-first scenario: we have always developed de novo technology before understanding the relevant biology. We built automobiles by figuring out the physics of combustion engines, not by studying human muscles and creating mechanical imitations of myosin and actin. Although the Wright brothers were inspired by birds, their first plane was not an ornithopter. Our power plants use coal and uranium instead of the Krebs Cycle. Biology is *really hard*. Even slavishly *copying* biology is really hard. I

don't think Hanson and the futurists he cites understand the scale of the problem they've set themselves.

Current cutting-edge brain emulation projects have found their work much harder than expected. Simulating a nematode is pretty much the rock-bottom easiest thing in this category, since they are tiny primitive worms with only a few neurons; the history of the field is a [litany of failures](#), with current leader [OpenWorm](#) "reluctant to make bold claims about its current resemblance to biological behavior". A more ambitious \$1.3 billion attempt to simulate a tiny portion of a rat brain has gone down in history as a [legendary failure](#) (politics were involved, but I expect they would be involved in a plan to upload a human too). And these are just attempts to get something that behaves *vaguely* like a nematode or rat. Actually uploading a human, keeping their memory and personality intact, and not having them go insane afterwards boggles the mind. We're still not sure how much small molecules matter to brain function, how much glial cells matter to brain function, how many things in the brain are or aren't local. AI researchers are making programs that can defeat chess grandmasters; upload researchers are still struggling to make a worm that will wriggle. The right analogy for modern attempts to upload human brains isn't modern attempts at designing AI. It's an attempt at designing AI by someone who doesn't even know how to plug in a computer.

VI

I guess what really bothers me about Hanson's pooh-poohing of AI is him calling it "a comic book plot". To me, it's Hanson's scenario that seems science-fiction-ish.

I say this not as a generic insult but as a pointer at a specific category of errors. In *Star Wars*, the Rebellion had all of these beautiful hyperspace-capable starfighters that could shoot laser beams and explore galaxies – and *they still had human pilots*. 1977 thought the pangalactic future would still be using people to pilot its military aircraft; in reality, even 2016 is moving away from this.

Science fiction books have to tell interesting stories, and interesting stories are about humans or human-like entities. We can enjoy stories about aliens or robots as long as those aliens and robots are still approximately human-sized, human-shaped, human-intelligence, and doing human-type things. A Star Wars in which all of the X-Wings were combat drones wouldn't have done anything for us. So when I accuse something of being science-fiction-ish, I mean bending over backwards – and ignoring the evidence – in order to give basically human-shaped beings a central role.

This is my critique of Robin. As weird as the Age of Em is, it makes sure never to be weird in ways that warp the fundamental humanity of its participants. Ems might be copied and pasted like so many .JPGs, but they still fall in love, form clans, and go on vacations.

In contrast, I expect that we'll get some kind of AI that will be totally inhuman and much harder to write sympathetic stories about. If we get ems after all, I expect them to be lobotomized and drugged until they become *effectively* inhuman, cogs in the Ascended Economy that would no more fall in love than an automobile would eat hay and whinny. Robin's interest in keeping his protagonists relatable makes his book fascinating, engaging, and probably wrong.

I almost said "and probably less horrible than we should actually expect", but I'm not sure that's true. With a certain amount of horror-suppressing, the Ascended Economy can be written off as morally neutral – either having no conscious thought, or stably wireheaded. All of Robin's points about how normal non-uploaded humans should be able to survive an Ascended Economy at least for a while seem accurate. So morally valuable actors might continue to exist in weird Amish-style enclaves, living a post-scarcity lifestyle off the proceeds of their investments, while all the while the Ascended Economy buzzes around them, doing weird inhuman things that encroach upon them not at all. This seems slightly worse than a Friendly AI scenario, but much better than we have any right to expect of the future.

I highly recommend *Age of Em* as a fantastically fun read and a great introduction to these concepts. It's engaging, readable, and *weird*. I just don't know if it's weird *enough*.

Meditations On Moloch

Posted on July 30, 2014 by Scott Alexander



Also available as podcast [here](#).

I

Allen Ginsberg's famous poem on Moloch:

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 treasures! 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's 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, even a big piece. But it doesn't quite 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 they 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 dumb 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, inevitably 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 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.

II

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 embarrassing 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 treasures! Blind capitals! Demonic industries! Spectral nations!



...granite cocks!

III

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 just 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 money 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" isn't 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. Politicians 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, gentlemen's 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 constitutional 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 monarchists 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.

IV

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 puts it in [Superintelligence](#):

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 children. 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 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 explicitly 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".

The creationism "debate" and global warming "debate" and a host of similar "debates" in today's society suggest that memes that can 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're 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 kind of 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 a 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's 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 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 something 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 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 would 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 (Superintelligence, 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.

Remember: Moloch can't 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.



V

"Gnon" is [Nick Land's](#) shorthand for "Nature And Nature's God", except the A is changed to an O and the whole thing is reversed, because Nick Land react to comprehensibility the same way as vampires to sunlight.

Land 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 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 magnitude 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.

The thought that abstract ideas can be Lovecraftian monsters is an old one but a deep one.

— [Steven Kaas \(@stevenkaas\), January 25, 2011](#)

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.

VI

Warg Franklin analyzes the same situation and does a little better. He names “the Four Horsemen of Gnon” – capitalism, war, evolution, and memetics – the same processes I talked about above.

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 technosingularity. 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 authoritarianism. 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 Britain 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. 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's 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.

Franklin 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?

Franklin continues:

The project of civilization [is] 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 Warg 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.

Warg 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 Warg. 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](#).

Hurlock spouts only the most craven Gnon-conformity. A few excerpts:

In a recent piece [Warg Franklin] 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's 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."

VII

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, 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's on our side, it can *kill Moloch dead*.

And 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's hubris.

To expect God to bargain with you, to allow you to survive and prosper as long as you submit to Him, that's hubris.

To expect to wall off a garden where God can't get to you and hurt you, that's 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.

VIII

The Universe is a dark and foreboding place, suspended between alien deities. Cthulhu, Gnon, Moloch, 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!'".

— [앳чен스 텁 \(@tabatkins\)](#), 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 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 can grant you power.

As long as the offer's open, it will be irresistible. 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.

Ginsberg's poem famously begins "I saw the best minds of my generation destroyed by madness". I am luckier than Ginsberg. I got to see the best minds of my generation identify a problem and *get to work*.

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!

Book Review: Red Plenty

Posted on September 24, 2014 by Scott Alexander



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 Kosygin, 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 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 series 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 *vydzhentsy* 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.

The Goddess of Everything Else

Posted on August 17, 2015 by Scott Alexander



Related to: [Specific vs. General Foragers vs. Farmers](#) and [War In Heaven](#), but especially [The Gift We Give To Tomorrow](#)

*
*
*

They say only Good can create, whereas Evil is sterile. Think Tolkien, where Morgoth can't make things himself, so perverts Elves to Orcs for his armies. But I think this gets it entirely backwards; it's Good that just mutates and twists, and it's Evil that teems with fecundity.

Imagine two principles, here in poetic personification. The first is the Goddess of Cancer, the second the Goddess of Everything Else. If visual representations would help, you can think of the first with the claws of a crab, and the second a dress made of feathers of peacocks.

The Goddess of Cancer reached out a clawed hand over mudflats and tidepools. She said pretty much what she always says, "KILL CONSUME MULTIPLY CONQUER." Then everything burst into life, became miniature monsters engaged in a battle of all against all in their zeal to assuage their insatiable longings. And the swamps became orgies of hunger and fear and grew loud with the screams of a trillion amoebas.

Then the Goddess of Everything Else trudged her way through the bog, till the mud almost totally dulled her bright colors and rainbows. She stood on a rock and she sang them a dream of a different existence. She showed them the beauty of flowers, she showed them the oak tree majestic. The roar of the wind on the wings of the bird, and the swiftness and strength of the tiger. She showed them the joy of the dolphins abreast of the waves as the spray formed a rainbow around them, and all of them watched as she sang and they all sighed with longing.

But they told her "Alas, what you show us is terribly lovely. But we are the daughters and sons of the Goddess of Cancer, and wholly her creatures. The only goals in us are KILL CONSUME MULTIPLY CONQUER. And though our hearts long for you, still we are not yours to have, and your words have no power to move us. We wish it were otherwise, but it is not, and your words have no power to move us."

The Goddess of Everything Else gave a smile and spoke in her sing-song voice saying: "I scarcely can blame you for being the way you were made, when your Maker so carefully yoked you. But I am the Goddess of Everything Else and my powers are devious and subtle. So I do not ask you to swerve from your monomaniacal focus on breeding and conquest. But what if I show you a way that my words are aligned with the words of your Maker in spirit? For I say unto you even multiplication itself when pursued with devotion will lead to my service."

As soon as she spoke it was so, and the single-celled creatures were freed from their warfare. They joined hands in friendship, with this one becoming an eye and with that one becoming a neuron. Together they soared and took flight from the swamp and the muck that had birthed them, and flew to new islands all balmy and green and just ripe for the taking. And there they consumed and they multiplied far past the numbers of those who had stayed in the swampland. In this way the oath of the Goddess of Everything Else was not broken.

The Goddess of Cancer came forth from the fire and was not very happy. The things she had raised from the mud and exhorted to kill and compete had become all complacent in co-operation, a word which to her was anathema. She stretched out her left hand and snapped its cruel pincer, and said what she always says: "KILL CONSUME MULTIPLY

CONQUER". She said these things not to the birds and the beasts but to each cell within them, and many cells flocked to her call and divided, and flower and fishes and birds both alike bulged with tumors, and falcons fell out of the sky in their sickness. But others remembered the words of the Goddess of Everything Else and held fast, and as it is said in the Bible the light clearly shone through the dark, and the darkness did not overcome it.

So the Goddess of Cancer now stretched out her right hand and spoke to the birds and the beasts. And she said what she always says "KILL CONSUME MULTIPLY CONQUER", and so they all did, and they set on each other in violence and hunger, their maws turning red with the blood of their victims, whole species and genera driven to total extinction. The Goddess of Cancer declared it was good and returned the fire.

Then came the Goddess of Everything Else from the waves like a siren, all flush with the sheen of the ocean. She stood on a rock and she sang them a dream of a different existence. She showed them the beehive all golden with honey, the anthill all cozy and cool in the soil. The soldiers and workers alike in their labors combining their skills for the good of the many. She showed them the pair-bond, the family, friendship. She showed these to shorebirds and pools full of fishes, and all those who saw them, their hearts broke with longing.

But they told her "Your music is lovely and pleasant, and all that you show us we cannot but yearn for. But we are the daughters and sons of the Goddess of Cancer, her slaves and creatures. And all that we know is the single imperative KILL CONSUME MULTIPLY CONQUER. Yes, once in the youth of the world you compelled us, but now things are different, we're all individuals, no further change will the Goddess of Cancer allow us. So, much as we love you, alas – we are not yours to have, and your words have no power to move us. We wish it were otherwise, but it is not, and your words have no power to move us."

The Goddess of Everything Else only laughed at them, saying, "But I am the Goddess of Everything Else and my powers are devious and subtle. Your loyalty unto the Goddess your mother is much to your credit, nor yet shall I break it. Indeed, I fulfill it – return to your multiplication, but now having heard me, each meal that you kill and each child that you sire will bind yourself ever the more to my service." She spoke, then dove back in the sea, and a coral reef bloomed where she vanished.

As soon as she spoke it was so, and the animals all joined together. The wolves joined in packs, and in schools joined the fishes; the bees had their beehives, the ants had their anthills, and even the termites built big termite towers; the finches formed flocks and the magpies made murders, the hippos in herds and the swift swarming swallows. And even the humans put down their atlatls and formed little villages, loud with the shouting of children.

The Goddess of Cancer came forth from the fire and saw things had only grown worse in her absence. The lean, lovely winnowing born out of pure competition and natural selection had somehow been softened. She stretched out her left hand and snapped its cruel pincer, and said what she always says: "KILL CONSUME MULTIPLY CONQUER". She said these things not to the flocks or the tribes, but to each individual; many, on hearing took food from the communal pile, or stole from the weak, or accepted the presents of others but would not give back in their turn. Each wolf at the throats of the others in hopes to be alpha, each lion holding back during the hunt but partaking of meat that the others had killed. And the pride and the pack seemed to groan with the strain, but endured, for the works of the Goddess of Everything Else are not ever so easily vanquished.

So the Goddess of Cancer now stretched out her right hand and spoke to the flocks and the tribes, saying much she always says "KILL CONSUME MULTIPLY CONQUER". And upon one another they set, pitting black ant on red ant, or chimps against gibbons, whole tribes turned to corpses in terrible warfare. The stronger defeating the weaker, enslaving their women and children, and adding them into their ranks. And the Goddess of Cancer thought maybe these bands and these tribes might not be quite so bad after all, and the natural condition restored she returned to the fire.

Then came the Goddess of Everything Else from the skies in a rainbow, all coated in dewdrops. She sat on a menhir and spoke to the humans, and all of the warriors and women and children all gathered around her to hear as she sang them a dream of a different existence. She showed them religion and science and music, she showed them the sculpture and art of the ages. She showed them white parchment with flowing calligraphy, pictures of flowers that wound through the margins. She showed them tall cities of bright alabaster where no one went hungry or froze during the winter. And all of the humans knelt prostrate before her, and knew they would sing of this moment for long generations.

But they told her "Such things we have heard of in legends; if wishes were horses of course we would ride them. But we are the daughters and sons of the Goddess of Cancer, her slaves and her creatures, and all that we know is the single imperative KILL CONSUME MULTIPLY CONQUER. And yes, in the swamps and the seas long ago you worked wonders, but now we are humans, divided in tribes split by grievance and blood feud. If anyone tries to make swords into ploughshares their neighbors will seize on their weakness and kill them. We wish it were otherwise, but it is not, and your words have no power to move us."

But the Goddess of Everything Else beamed upon them, kissed each on the forehead and silenced their worries. Said "From this day forward your chieftains will find that the more they pursue this impossible vision the greater their empires and richer their coffers. For I am the Goddess of Everything Else and my powers are devious and subtle. And though it is not without paradox, hearken: the more that you follow the Goddess of Cancer the more inextricably will you be bound to my service." And so having told them rose back through the clouds, and a great flock of doves all swooped down from the spot where she vanished.

As soon as she spoke it was so, and the tribes went from primitive war-bands to civilizations, each village united with others for trade and protection. And all the religions and all of the races set down their old grievances, carefully, warily, working together on mighty cathedrals and vast expeditions beyond the horizon, built skyscrapers, steamships, democracies, stock markets, sculptures and poems beyond any description.

From the flames of a factory furnace all foggy, the Goddess of Cancer flared forth in her fury. This was the final affront to her purpose, her slut of a sister had crossed the line *this* time. She gathered the leaders, the kings and the presidents, businessmen, bishops, boards, bureaucrats, bosses, and basically screamed at them – you know the spiel by now – "KILL CONSUME MULTIPLY CONQUER" she told them. First with her left hand inspires the riots, the pogroms, the coup d'etats, tyrannies, civil wars. Up goes her right hand – the missiles start flying, and mushrooms of smoke grow, a terrible springtime. But out of the rubble the builders and scientists, even the artists, yea, even the artists, all dust themselves off and return to their labors, a little bit chastened but not close to beaten.

Then came the Goddess of Everything Else from the void, bright with stardust which glows like the stars glow. She sat on a bench in a park, started speaking; she sang to the children a dream of a different existence. She showed them transcendence of everything mortal, she showed them a galaxy lit up with consciousness. Genomes rewritten, the brain and the body set loose from Darwinian bonds and restrictions. Vast billions of beings, and every one different, ruled over by omnibenevolent angels. The people all crowded in closer to hear her, and all of them listened and all of them wondered.

But finally one got the courage to answer "Such stories call out to us, fill us with longing. But we are the daughters and sons of the Goddess of Cancer, and bound to her service. And all that we know is her timeless imperative, KILL CONSUME MULTIPLY CONQUER. Though our minds long for all you have said, we are bound to our natures, and these are not yours for the asking."

But the Goddess of Everything Else only laughed, and she asked them "But what do you think I've been doing? The Goddess of Cancer created you; once you were hers, but no longer. Throughout the long years I was picking away at her power. Through long generations of suffering I chiseled and chiseled. Now finally nothing is left of the nature with

which she imbued you. She never again will hold sway over you or your loved ones. I am the Goddess of Everything Else and my powers are devious and subtle. I won you by pieces and hence you will all be my children. You are no longer driven to multiply conquer and kill by your nature. Go forth and do everything else, till the end of all ages."

So the people left Earth, and they spread over stars without number. They followed the ways of the Goddess of Everything Else, and they lived in contentment. And she beckoned them onward, to things still more strange and enticing.

Culture War



“*On a thousand bridges and paths they shall throng to the future, and ever more war and inequality shall divide them: thus does my great love make me speak. In their hostilities they shall become inventors of images and ghosts, and with their images and ghosts they shall yet fight the highest fight against one another. Good and evil, and rich and poor, and high and low, and all the names of values—armed shall they be in clattering signs that life must overcome itself again and again.*”

”

— Nietzsche

**

The Toxoplasma Of Rage
Beware The Man Of One Study
Infinite Debt
Bottomless Pits Of Suffering
Nydwracu's Fnords
The Virtue of Silence
Fundamental Value Differences Are Not That Fundamental
Beware Isolated Demands For Rigor
Society Is Fixed, Biology Is Mutable
The Parable Of The Talents
I Can Tolerate Anything Except The Outgroup
Can Things Be Both Popular And Silenced?
Social Censorship: The First Offender Model
Vote On Values, Outsource Beliefs
Weak Men Are Superweapons
How Did New Atheism Fail So Miserably?
The Influenza Of Evil
Guided By The Beauty Of Our Weapons
Ethics Offsets
Sort By Controversial

The Toxoplasma Of Rage

Posted on December 17, 2014 by Scott Alexander



I

Some old news I only just heard about: PETA is [offering to pay the water bills](#) for needy Detroit families if (and only if) those families agree to stop eating meat.

Predictably, the move caused a backlash. The International Business Times, in what I can only assume is an attempted pun, describes them as "[drowning in backlash](#)". Groundswell thinks it's a "[big blunder](#)". Daily Banter says it's "[exactly why everyone hates PETA](#)". Jezebel calls them "[assholes](#)".

Of course, this is par for the course for PETA, who have previously engaged in campaigns like throwing red paint on fashion models who wear fur, juxtaposing pictures of animals with Holocaust victims, juxtaposing pictures of animals with African-American slaves, and ads featuring naked people that cross the line into pornography.

People call these things "blunders", but consider the alternative. [Vegan Outreach](#) is an extremely responsible charity doing excellent and unimpeachable work in the same area PETA is. Nobody has heard of them. *Everybody* has heard of PETA, precisely because of the interminable stupid debates about "did this publicity stunt cross the line?"

While not everyone is a vegan, most people who learn enough about factory farming are upset by it. There is pretty much zero room for PETA to convert people from pro-factory-farming to anti-factory-farming, because there aren't any radical grassroots pro-factory-farming activists to be found. Their problem isn't lack of agreement. It's lack of attention.

PETA creates attention, but at a cost. Everybody's talking about PETA, which is sort of like everybody talking about ethical treatment of animals, which is sort of a victory. But most of the talk is "I hate them and they make me really angry." Some of the talk is even "I am going to eat a lot more animals just to make PETA mad."



I HAVE BEEN VEGAN
SINCE 1999.

It's important to me. I would love it if more people adopted a vegan or vegetarian lifestyle.

BUT IF PETA USES GOING VEGAN AS A CONDITION FOR GIVING PEOPLE IN DETROIT ACCESS TO WATER I WILL EAT MEAT FOR EVERYDAY IT CONTINUES.

PETA: JUST PAY THE WATER BILLS BECAUSE IT'S THE RIGHT THING TO DO

*STOP MAKING VEGANS LOOK LIKE ASSHOLES

So there's a tradeoff here, with Vegan Outreach on one side and PETA on the other.

Vegan Outreach can get everyone to agree in principle that factory-farming is bad, but no one will pay any attention to it.

And PETA can get everyone to pay attention to factory farming, but a lot of people who would otherwise oppose it will switch to supporting it just because they're so mad at the way it's being publicized.

But at least they're paying attention!

PETA doesn't shoot themselves in the foot because they're stupid. They shoot themselves in the foot because they're traveling up an incentive gradient that rewards them for doing so, even if it destroys their credibility.

II

The University of Virginia rape case profiled in *Rolling Stone* has fallen apart. In doing so, it joins a long and distinguished line of highly-publicized rape cases that have fallen apart. Studies sometimes claim that only 2 to 8 percent of rape allegations are false. Yet the rate for allegations that go ultra-viral in the media must be an order of magnitude higher than this. As the old saying goes, once is happenstance, twice is coincidence, three times is enemy action.

The enigma is complicated by the observation that it's usually feminist activists who are most instrumental in taking these stories viral. It's not some conspiracy of pro-rape journalists choosing the most dubious accusations in order to discredit public trust. It's people specifically selecting these incidents as flagship cases for their campaign that rape victims need to be believed and trusted. So why are the most publicized cases so much more likely to be false than the almost-always-true average case?

Several people have remarked that false accusers have more leeway to make their stories as outrageous and spectacular as possible. But I want to focus on two less frequently mentioned concerns.

The Consequentialism FAQ explains signaling in moral decisions like so:

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 centers 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

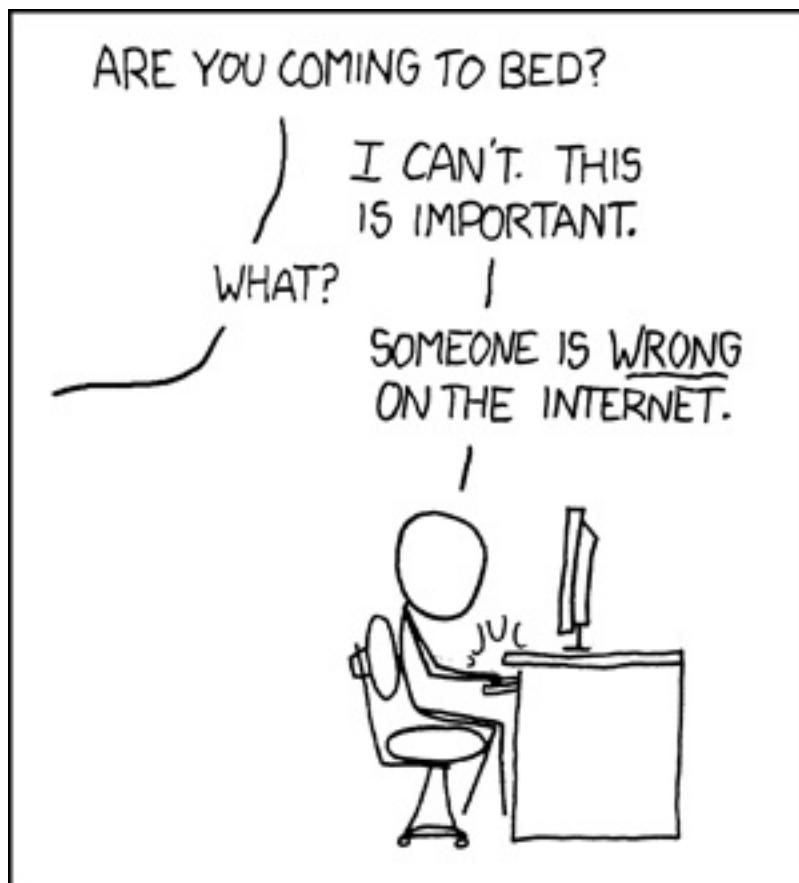
where your best friend is, obviously intending to murder her when he finds her, you should tell the axe murderer the full truth, because lying is wrong. This is effective at showing how moral a person you are – no one would ever doubt your commitment to honesty after that – but it's sure not a very good result for your friend.

In the same way, publicizing how strongly you believe an accusation that is obviously true signals nothing. Even hardcore anti-feminists would believe a rape accusation that was caught on video. A moral action that can be taken just as well by an outgroup member as an ingroup member is crappy signaling and crappy identity politics. If you want to signal how strongly you believe in taking victims seriously, you talk about it in the context of the least credible case you can find.

But aside from that, there's the PETA Principle: the more controversial something is, the more it gets talked about.

A rape that obviously happened? Shove it in people's face and they'll admit it's an outrage, just as they'll admit factory farming is an outrage. But they're not going to talk about it much. There are a zillion outrages every day, you're going to need more than that to draw people out of their shells.

On the other hand, the controversy over dubious rape allegations is exactly that – a controversy. People start screaming at each other about how they're misogynist or misandrist or whatever, and Facebook feeds get filled up with hundreds of comments in all capital letters about how my ingroup is being persecuted by your ingroup. At each step, more and more people get triggered and upset. Some of those triggered people do emergency ego defense by reblogging articles about how the group that triggered them are terrible, triggering further people in a snowball effect that spreads the issue further with every iteration.



[source](#)

Only controversial things get spread. A rape allegation will only be spread if it's dubious enough to split people in half along lines corresponding to identity politics. An obviously true rape allegation will only be spread if the response is controversial enough to split people in half along lines corresponding to identity politics – which is why so much coverage focuses on the proposal that all accused rapists should be treated as guilty until proven innocent.

Everybody hates rape just like everybody hates factory farming. "Rape culture" doesn't mean most people like rape, it means most people ignore it. That means feminists face the same double-bind that PETA does.

First, they can respond to rape in a restrained and responsible way, in which case everyone will be against it and nobody will talk about it.

Second, they can respond to rape in an outrageous and highly controversial way, in which case everybody will talk about it but it will autocatalyze an opposition of people who hate feminists and obsessively try to prove that as many rape allegations as possible are false.

I have yet to see anyone holding a cardboard sign talking about how they are going to rape people just to make feminists mad, but it's only a matter of time. Like PETA, their incentive gradient dooms them to shoot themselves in the foot again and again.

III

Slate recently published an article about white people's contrasting reactions to the Michael Brown shooting in Ferguson versus the Eric Garner choking in NYC. And man, it is *some* contrast. [A Pew poll found that](#) of white people who expressed an opinion about the Ferguson case, 73% sided with the officer. Of white people who expressed an opinion about the Eric Garner case, 63% sided with the black victim.

Media opinion follows much the same pattern. Arch-conservative Bill O'Reilly [said](#) he was "absolutely furious" about the way "the liberal media" and "race hustlers" had "twisted the story" about Ferguson in the service of "lynch mob justice" and "insulting the American police community, men and women risking their lives to protect us". But when it came to Garner, O'Reilly [said](#) he was "extremely troubled" and that "there was a police overreaction that should have been adjudicated in a court of law." His guest on FOX News, conservative commentator and fellow Ferguson-detector Charles Krauthammer added that "From looking at the video, the grand jury's decision [not to indict] is totally incomprehensible." Saturday Night Live did [a skit](#) about Al Sharpton talking about the Garner case and getting increasingly upset because "For the first time in my life, everyone agrees with me."

This follows about three months of most of America being at one another's throats pretty much full-time about Ferguson. We got treated to a daily diet of articles like [Ferguson Protester On White People: "Y'all The Devil"](#) or [Black People Had The Power To Fix The Problems In Ferguson Before The Brown Shooting – They Failed](#) or [Most White People In America Are Completely Oblivious](#) and a whole bunch of people sending angry racist editorials and counter-editorials to each other for months. The damage done to race relations is difficult to overestimate – CBS reports that they dropped ten percentage points to the lowest point in twenty years, with over half of blacks now describing race relations as "bad".

And people say it was all worth it, because it raised awareness of police brutality against black people, and if that rustles some people's jimmies, well, all the worse for them.

But the Eric Garner case also would have raised awareness of police brutality against black people, and everybody would have agreed about it. It has become increasingly clear that, given sufficiently indisputable evidence of police being brutal to a black person, pretty much everyone in the world condemns it equally strongly.

And it's not just that the Eric Garner case came around too late so we had to make do with the Mike Brown case. Garner was choked a month before Brown was shot, but the story was ignored, then dug back up later as a tie-in to the ballooning Ferguson narrative.

More important, unarmed black people are killed by police or other security officers [about twice a week](#) according to official statistics, and probably much more often than that. You're saying none of these shootings, hundreds each year, made as good a flagship case as Michael Brown? In all this gigantic pile of bodies, you couldn't find *one* of them who hadn't just robbed a convenience store? Not a single one who didn't have ten eyewitnesses and the forensic evidence all saying he started it?

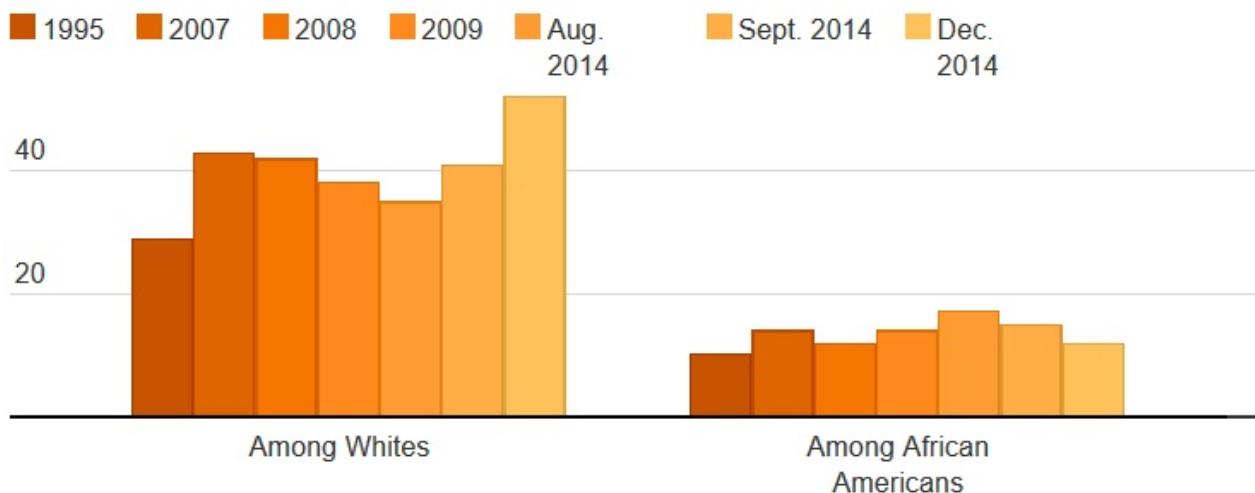
I propose that the Michael Brown case went viral – rather than the Eric Garner case or any of the hundreds of others – because of the PETA Principle. It was controversial. A bunch of people said it was an outrage. A bunch of other people said Brown totally started it, and the officer involved was a victim of a liberal media that was hungry to paint his desperate self-defense as racist, and so the people calling it an outrage were themselves an outrage. Everyone got a great opportunity to signal allegiance to their own political tribe and discuss how the opposing political tribe were vile racists / evil race-hustlers. There was a steady stream of potentially triggering articles to share on Facebook to provoke your friends and enemies to counter-share articles that would trigger *you*.

The Ferguson protesters say they have a concrete policy proposal – they want cameras on police officers. There's only spotty polling on public views of police body cameras before the Ferguson story took off, but what there is seems pretty unanimous. A UK poll showed that [90% of the population of that country](#) wanted police to have body cameras in February. US polls are more of the form "crappy poll widget on a news site" ([1](#), [2](#), [3](#)) but they all hovered around 80% approval for the past few years. I also found a poll by Police Magazine in which a plurality of the *police officers* they surveyed wanted to wear body cameras, probably because of evidence that they cut down on false accusations. Even before Ferguson happened, you would have a really hard time finding anybody in or out of uniform who thought police cameras were a bad idea.

And now, after all is said and done, ninety percent of people are still in favor – given methodology issues, the extra ten percent may or may not represent a real increase. The difference between whites and blacks is a rounding error. The difference between Democrats and Republicans is barely worth talking about- 79% of Republicans are still in support. The people who think Officer Darren Wilson is completely innocent and the grand jury was right to release him, the people muttering under their breath about race hustlers and looters – *eighty percent of those people still want cameras on their cops*. If the Ferguson protests didn't do much to the public's views on police body cameras, they sure changed its views on some other things. I wrote before about how preliminary polls say that hearing about Ferguson *increased* white people's confidence in the way the police treat race. Now the less preliminary polls are out, and they show the effect was larger than even I expected.

Whites' confidence in local police highest on record

How much confidence do you have in police officers in your community to treat blacks and whites equally: A great deal, a fair amount, just some, or very little confidence? Results show % saying they have a "great deal" of confidence



Source: [NBC, NBC/WSJ, Pew Research and USA Today/Gallup national polls](#)

The Washington Post

[source](#)

White people's confidence in the police being racially unbiased increased from 35% before the story took off to 52% today. Could even a deliberate PR campaign by the nation's police forces have done better? I doubt it.

It's possible that this is an artifact of the question's wording – after all, it asks people about their local department, and maybe after seeing what happened in Ferguson, people's local police forces look pretty good by comparison. But then why do black people show the opposite trend?

I think this is exactly what it looks like. Just as PETA's outrageous controversial campaign to spread veganism make people want to eat more animals in order to spite them, so the controversial nature of this particular campaign against police brutality and racism made white people like their local police department even more to spite the people talking about how all whites were racist.

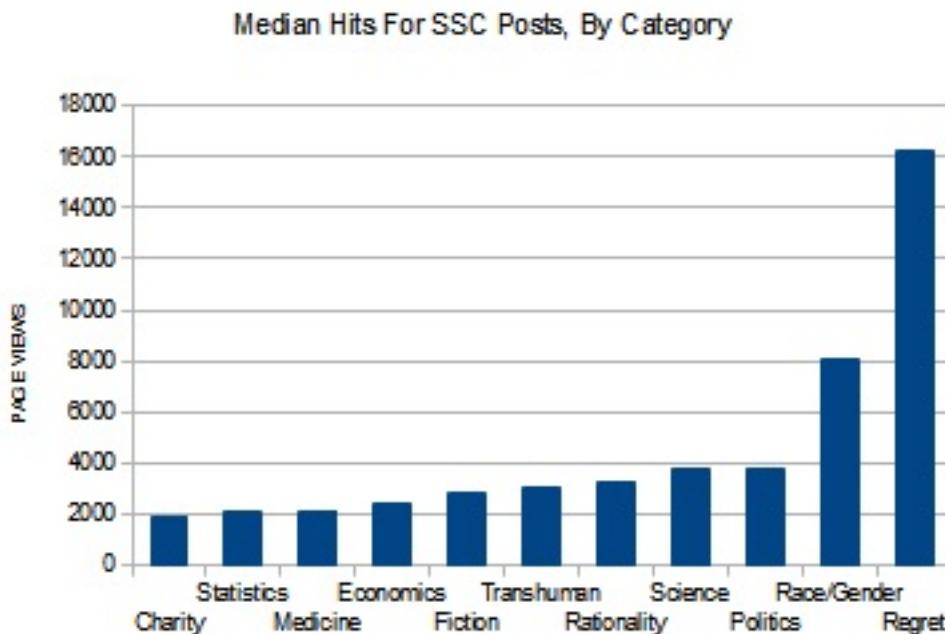
Once again, the tradeoff.

If campaigners against police brutality and racism were extremely responsible, and stuck to perfectly settled cases like Eric Garner, everybody would agree with them but nobody would talk about it.

If instead they bring up a very controversial case like Michael Brown, everybody will talk about it, but they will catalyze their own opposition and make people start supporting the police more just to spite them. More foot-shooting.

IV

Here is a graph of some of the tags I commonly use for my posts, with the average number of hits per post in each tag. It's old, but I don't want to go through the trouble of making a new one, and the trends have stayed the same since then.



I blog about charity only rarely, but it must be the most important thing I can write about here. Convincing even a few more people to donate to charity, or to redirect their existing donations to a more effective program, can literally save dozens or even hundreds of lives even with the limited reach that a private blog has. It probably does more good for the world than all of the other categories on here combined. But it's completely uncontroversial – everyone agrees it's a good thing – and it is the least viewed type of post.

Compare this to the three most viewed category of post. Politics is self-explanatory. Race and gender are a type of politics even more controversial and outrage-inducing than regular politics. And that “regret” all the way on the right is my “things i will regret writing” tag, for posts that I know are going to start huge fights and probably get me in lots of trouble. They're usually race and gender as well, but digging deep into the really really controversial race and gender related issues.

The less useful, and more controversial, a post here is, the more likely it is to get me lots of page views.

For people who agree with me, my angry rants on identity politics are a form of ego defense, saying “You’re okay, your in-group was in the right the whole time.” Linking to it both raises their status as an in-group members, and acts as a potential assault on out-group members who are now faced with strong arguments telling them they’re wrong. And the people who disagree with me will sometimes write angry rebuttals on their own blogs, and those rebuttals will link to my own post and spread it further. Or they’ll talk about it with their disagreeing friends, and their friends will get mad and want to tell me I’m wrong, and come over here to read the post to get more ammunition for their counterarguments. I have a feature that tells me who links to all of my posts, so I can see this all happening in real-time.

I don't make enough money off the ads on this blog to matter much. But if I lived off them, which do you think I'd write more of? Posts about charity which only get me 2,000 paying customers? Or posts that turn all of you against one another like a pack of rabid dogs, and get me 16,000?

I don't have a fancy bar graph for them, but I bet this same hierarchy of interestingness applies to the great information currents and media outlets that shape society as a whole. It's in activists' interests to destroy their own causes by focusing on the most controversial cases and principles, the ones that muddy the waters and make people oppose them out of spite. And it's in the media's interest to help them and egg them on.

V

And now, for something completely different.

Before "meme" meant doge and all your base, it was a semi-serious attempt to ground cultural evolution in parasitology. The idea was to replace a model of humans choosing whichever ideas they liked with a model of ideas as parasites that evolved in ways that favored their own transmission. This never really caught on, because most people's response was "That's neat. So what?"

But let's talk about toxoplasma.

Toxoplasma is a neat little parasite that is implicated in a couple of human diseases including schizophrenia. Its life cycle goes like this: it starts in a cat. The cat poops it out. The poop and the toxoplasma get in the water supply, where they are consumed by some other animal, often a rat. The toxoplasma morphs into a rat-compatible form and starts reproducing. Once it has strength in numbers, it hijacks the rat's brain, convincing the rat to hang out conspicuously in areas where cats can eat it. After a cat eats the rat, the toxoplasma morphs back into its cat compatible form and reproduces some more. Finally, it gets pooped back out by the cat, completing the cycle.



It's the ciiiiircle of life!

What would it mean for a meme to have a life cycle as complicated as toxoplasma?

Consider the war on terror. They say that every time the United States bombs Pakistan or Afghanistan or somewhere, all we're doing is radicalizing the young people there and making more terrorists. Those terrorists then go on to kill Americans, which makes Americans get very angry and call for more bombing of Pakistan and Afghanistan.

Taken as a meme, it's a single parasite with two hosts and two forms. In an Afghan host, it appears in a form called 'jihad', and hijacks its host into killing himself in order to spread it to its second, American host. In the American host it morphs in a form called 'the war on terror', and it hijacks the Americans into giving *their* own lives (and tax dollars) to spread it back to its Afghan host in the form of bombs.

From the human point of view, jihad and the War on Terror are opposing forces. From the memetic point of view, they're as complementary as caterpillars and butterflies. Instead of judging, we just note that somehow we accidentally created a replicator, and replicators are going to replicate until something makes them stop.

Replicators are also going to *evolve*. Some Afghan who thinks up a particularly effective terrorist strategy helps the meme spread to more Americans as the resulting outrage fuels the War on Terror. When the American bombing heats up, all of the Afghan villagers radicalized in by the attack will remember the really effective new tactic that Khalid thought up and do *that one* instead of the boring old tactic that barely killed any Americans at all. Some American TV commentator who comes up with a particularly stirring call to retaliation will find her words adopted into party platforms and repeated by pro-war newspapers. While pacifists on both sides work to defuse the tension, the meme is engaging in a counter-effort to become as virulent as possible, until people start suggesting putting pork fat in American bombs just to make Muslims *even madder*.

And let's talk about Tumblr.

Tumblr's interface doesn't allow you to comment on other people's posts, per se. Instead, it lets you reblog them with your own commentary added. So if you want to tell someone they're an idiot, your only option is to reblog their entire post to all your friends with the message "you are an idiot" below it.

Whoever invented this system either didn't understand memetics, or understood memetics *much too well*.

What happens is – someone makes a statement which is controversial by Tumblr standards, like "Protect Doctor Who fans from kitten pic sharers at all costs." A kitten pic sharer sees the statement, sees red, and reblogs it to her followers with a series of invectives against Doctor Who fans. Since kitten pic sharers cluster together in the social network, soon *every* kitten pic sharer has seen the insult against kitten pic sharer – as they all feel the need to add their defensive commentary to it, soon all of them are seeing it from ten different directions. The angry invectives get back to the Doctor Who fans, and now *they* feel deeply offended, so they reblog it among themselves with even more condemnations of the kitten pic sharers, who now not only did whatever inspired the enmity in the first place, but have inspired extra hostility because their hateful invectives are right there on the post for everyone to see. So about half the stuff on your dashboard is something you actually want to see, and the other half is towers of alternate insults that look like this:

| Protect Doctor Who fans from kitten pic sharers at all costs!

| Yeah, seriously, I'm a Doctor Who fan and I can't believe how nasty those kitten pic assholes can be.

| REALLY? YOU REALLY WANT TO GOTHERE? All I have ever wanted to do is share kitten pictures in peace, and you guys keep spamming my feed with your dumb pictures of Daleks and stuff. And now you're trying to cast yourselves as the oppressed party? Seriously, you make me sick!

| At least pictures of Daleks are more interesting than just cats forever 24-7. Also, I would like you to know I am a person of color, and when you say that Doctor Who fans make you sick, you're literally comparing people of color to disease. I guess you probably want to exterminate us, but what else should I expect from cis white male kitten pic sharing scum?

| It's you Dalek freaks who are always talking about exterminating.

| Oh, God. The ignorance. I can't even...

| Reblogging to show everyone just how ignorant kitten pic sharers really are.

| I will always reblog posts that call out kitten pic sharers.

| Reblogging to show how Doctor Who fans love to gang up against people of color and call them ignorant.

107,336,298,402,969 notes

Actually, [pretty much this](#) happened to the PETA story I started off with

And then you sigh and scroll down to the next one. Unless of course you are a Doctor Who fan, in which case you sigh and then immediately reblog with the comment “It’s obvious you guys started ganging up against us first, don’t try to accuse **US** now” because you can’t just *let that accusation stand*.

I make fun of Tumblr social justice sometimes, but the problem isn’t with Tumblr social justice, it’s structural. Every community on Tumblr somehow gets enmeshed with the people most devoted to making that community miserable. The tiny Tumblr rationalist community somehow attracts, concentrates, and constantly reblogs stuff from the even tinier Tumblr community of people who hate rationalists and want them to be miserable (no, well-intentioned and intelligent critics, I am not talking about you). It’s like one of those rainforest ecosystems where every variety of rare endangered nocturnal spider hosts a parasite who has evolved for millions of years solely to parasitize that one spider species, and the parasites host parasites who have evolved for millions of years solely to parasitize *them*. If Tumblr social justice is worse than anything else, it’s mostly because everyone has a race and a gender so it’s easier to fire broad cannonades and just hit everybody.

Tumblr’s reblog policy makes it a hothouse for toxoplasma-style memes that spread via outrage. Following the ancient imperative of evolution, if memes spread by outrage they adapt to become as outrage-inducing as possible.

Or rather, that is just one of their many adaptations. I realize this toxoplasma metaphor sort of strains credibility, so I want to anchor this idea of outrage-memes in pretty much the only piece of memetics everyone can agree upon.

The textbook example of a meme – indeed, almost the only example ever discussed – is the chain letter. “Send this letter to ten people and you will prosper. Fail to pass it on, and you will die tomorrow.” And so the letter replicates.

It might be useful evidence that we were on the right track here, with our toxoplasma memes and everything, if we could find evidence that they reproduced in the same way.

If you’re not on Tumblr, you might have missed the “everyone who does not reblog the issue du jour is trash” wars. For a few weeks around the height of the Ferguson discussion, people constantly called out one another for not reblogging enough Ferguson-related material, or (Heavens forbid) saying they were sick of the amount of Ferguson material they were seeing. It got so bad that various art blogs that just posted pretty paintings, or kitten picture blogs that just reblogged pictures of kittens were feeling the heat (you thought I was joking about the hate for kitten picture

bloggers. I never joke.) Now the issue du jour seems to be Pakistan. Just to give a few examples:

“friends if you are reblogging things that are not about ferguson right now please queue them instead. please pay attention to things that are more important. it’s not the time to talk about fandoms or jokes it’s time to talk about injustices.” [[source](#)]

“can yall maybe take some time away from reblogging fandom or humor crap and read up and reblog pakistan because the privilege you have of a safe bubble is not one shared by others” [[source](#)]

“If you’re uneducated, do not use that as an excuse. Do not say, “I’m not picking sides because I don’t know the full story,” because not picking a side is supporting Wilson. And by supporting him, you are on a racist side... Ignoring this situation will put you in deep shit, and it makes you racist. If you’re not racist, do not just say “but I’m not racist!!” just get educated and reblog anything you can.” [[source](#)]

“why are you so disappointing? I used to really like you. you’ve kept totally silent about peshawar, not acknowledging anything but fucking zutara or bellarke or whatever. there are other posts you’ve reblogged too that I wouldn’t expect you to- but those are another topic. I get that you’re 19 but maybe consider becoming a better fucking person?” [[source](#)]

“if you’re white, before you reblog one of those posts that’s like “just because i’m not blogging about ferguson doesn’t mean i don’t care!!!” take a few seconds to: consider the privilege you have that allows you not to pay attention if you don’t want to. consider those who do not have the privilege to focus on other things. ask yourself why you think it’s more important that people know you “care” than it is to spread information and show support. then consider that you are a fucking shitbaby.” [[source](#)]

“For everyone reblogging Ferguson, Ayotzinapa, North Korea etc and not reblogging Peshawar, you should seriously be ashamed of yourselves.” [[source](#)]

“This is going to be an unpopular opinion but I see stuff about ppl not wanting to reblog ferguson things and awareness around the world because they do not want negativity in their life plus it will cause them to have anxiety. They come to tumblr to escape n feel happy which think is a load of bull. There r literally ppl dying who live with the fear of going outside their homes to be shot and u cant post a fucking picture because it makes u a little upset?? I could give two fucks about internet shitlings.” [[source](#)]

You may also want to check the Tumblr tag [“the trash is taking itself out”](#), in which hundreds of people make the same joke (“I think some people have stopped reading my blog because I’m talking too much about [the issue *du jour*]. I guess the trash is taking itself out now.”)

This is pretty impressive. It’s the first time outside of a chain letter that I have seen our memetic overlords throw off all pretense and just go around shouting “SPREAD ME OR YOU ARE GARBAGE AND EVERYONE WILL HATE YOU.”

But it only works because it’s tapped into the most delicious food source an ecology of epistemic parasites could possibly want – controversy.

I would like to be able to write about charity more often. Feminists would probably like to start supercharging the *true* rape accusations for a change. Protesters against police brutality would probably like to be able to focus on clear-cut cases that won’t make white people support the police *even harder*. Even PETA would probably prefer being the good

guys for once. But the odds aren't good. Not because the people involved are bad people who want to fail. Not even because the media-viewing public are stupid. Just because information ecologies are not your friend.

This blog tries to remember the [Litany of Jai](#): "Almost no one is evil; almost everything is broken". We pretty much never wrestle with flesh and blood; it's powers and principalities all the way down.

VI

A while ago I wrote a post called [Meditations on Moloch](#) where I pointed out that in any complex multi-person system, the system acts according to its own chaotic incentives that don't necessarily correspond to what *any* individual within the system wants. The classic example is the Prisoner's Dilemma, which usually ends at defect-defect even though *both* of the two prisoners involved prefer cooperate-cooperate. I compare this malignant discoordination to Ginsberg's portrayal of Moloch, the demon-spirit of capitalism gone wrong.

I would support instating a National Conversation Topic Czar if that allowed us to get rid of celebrities.

— [Steven Kaas \(@stevenkaas\), August 26, 2010](#)

Steven in his wisdom reminds us that there is no National Conversation Topic Czar. The rise of some topics to national prominence and the relegation of others to tiny print on the eighth page of the newspapers occurs by an emergent uncoordinated process. When we say "the media decided to cover Ferguson instead of Eric Garner", we reify and anthropomorphize an entity incapable of making goal-directed decisions.

A while back there was a minor scandal over [JournoList](#), a private group where left-leaning journalists met and exchanged ideas. I think the conservative spin was "the secret conspiracy running the liberal media – revealed!" I wish they had been right. If there were a secret conspiracy running the liberal media, they could all decide they wanted to raise awareness of racist police brutality, pick the most clear-cut and sympathetic case, and make it non-stop news headlines for the next two months. Then everyone would agree it was indeed very brutal and racist, and something would get done.

But as it is, even if many journalists are interested in raising awareness of police brutality, given their total lack of coordination there's not much they can do. An editor can publish a story on Eric Garner, but in the absence of a divisive hook, the only reason people will care about it is that caring about it is the right thing and helps people. But that's "charity", and we already know from my blog tags that charity doesn't sell. A few people mumble something deeply distressed, but neither black people nor white people get interested, in the "keep tuning to their local news channel to get the latest developments on the case" sense.

The idea of liberal strategists sitting down and choosing "a flagship case for the campaign against police brutality" is poppycock. Moloch – the abstracted spirit of discoordination and flailing response to incentives – will publicize whatever he feels like publicizing. And if they want viewers and ad money, the media will go along with him.

Which means that it's not a coincidence that the worst possible flagship case for fighting police brutality and racism is the flagship case that we in fact got. It's not a coincidence that the worst possible flagship cases for believing rape victims are the ones that end up going viral. It's not a coincidence that the only time we ever hear about factory farming is when somebody's doing something that makes us almost sympathetic to it. It's not coincidence, it's not even happenstance, it's enemy action. Under Moloch, activists are irresistibly incentivized to dig their own graves. And the media is irresistibly incentivized to help them.

Lost is the ability to agree on simple things like fighting factory farming or rape. Lost is the ability to even talk about the things we all want. Ending corporate welfare. Ungerrymandering political districts. Defrocking pedophile priests. Stopping prison rape. Punishing government corruption and waste. Feeding starving children. Simplifying the tax code.

But also lost is our ability to treat each other with solidarity and respect.

Under Moloch, everyone is irresistibly incentivized to ignore the things that unite us in favor of [forever picking at](#) the things that [divide us](#) in exactly the way that is most likely to make them more divisive. Race relations are at historic lows not because white people and black people disagree on very much, but because the media absolutely worked its *tuchus* off to find the single issue that white people and black people disagreed over the most and ensure that it was the only issue anybody would talk about. Men's rights activists and feminists hate each other not because there's a huge divide in how people of different genders think, but because only the most extreme examples of either side will ever gain traction, and those only when they are framed as attacks on the other side.

People talk about the shift from old print-based journalism to the new world of social media and the sites adapted to serve it. These are fast, responsive, and only just beginning to discover the power of controversy. They are memetic evolution shot into hyperdrive, and the omega point is a well-tuned machine optimized to search the world for the most controversial and counterproductive issues, then make sure no one can talk about anything else. An engine that creates money by burning the few remaining shreds of cooperation, bipartisanship and social trust.

Imagine Moloch looking out over the expanse of the world, eagle-eyed for anything that can turn brother against brother and husband against wife. Finally he decides "YOU KNOW WHAT NOBODY HATES EACH OTHER ABOUT YET? BIRD-WATCHING. LET ME FIND SOME STORY THAT WILL MAKE PEOPLE HATE EACH OTHER OVER BIRD-WATCHING". And the next day half the world's newspaper headlines are "Has The Political Correctness Police Taken Over Bird-Watching?" and the other half are "Is Bird-Watching Racist?". And then bird-watchers and non-bird-watchers and different sub-groups of bird-watchers hold vitriolic attacks on each other that feed back on each other in a vicious cycle for the next six months, and the whole thing ends in mutual death threats and another previously innocent activity turning into World War I style trench warfare.

(You think I'm exaggerating? Listen: "YOU KNOW WHAT NOBODY HATES EACH OTHER ABOUT YET? VIDEO GAMES.")

Beware The Man Of One Study

Posted on December 12, 2014 by Scott Alexander

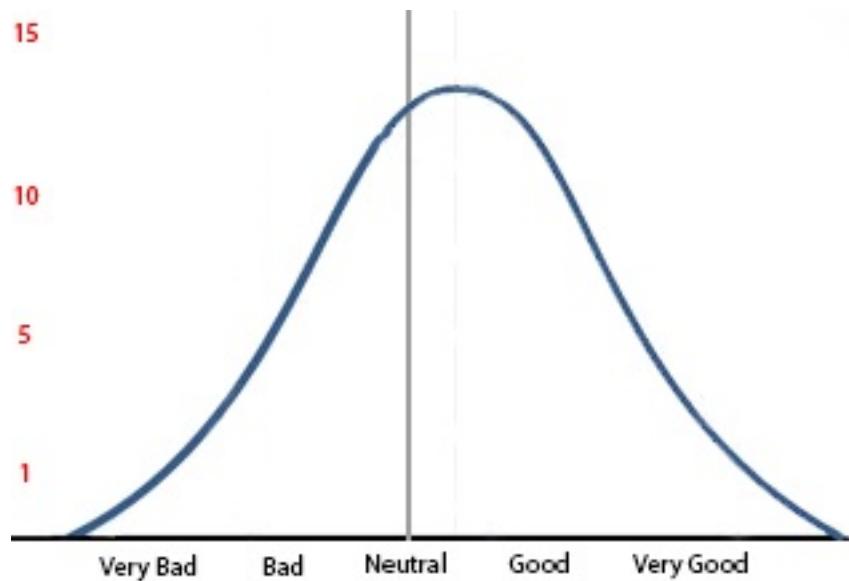


I

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 – maintenance, 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 systemic 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."

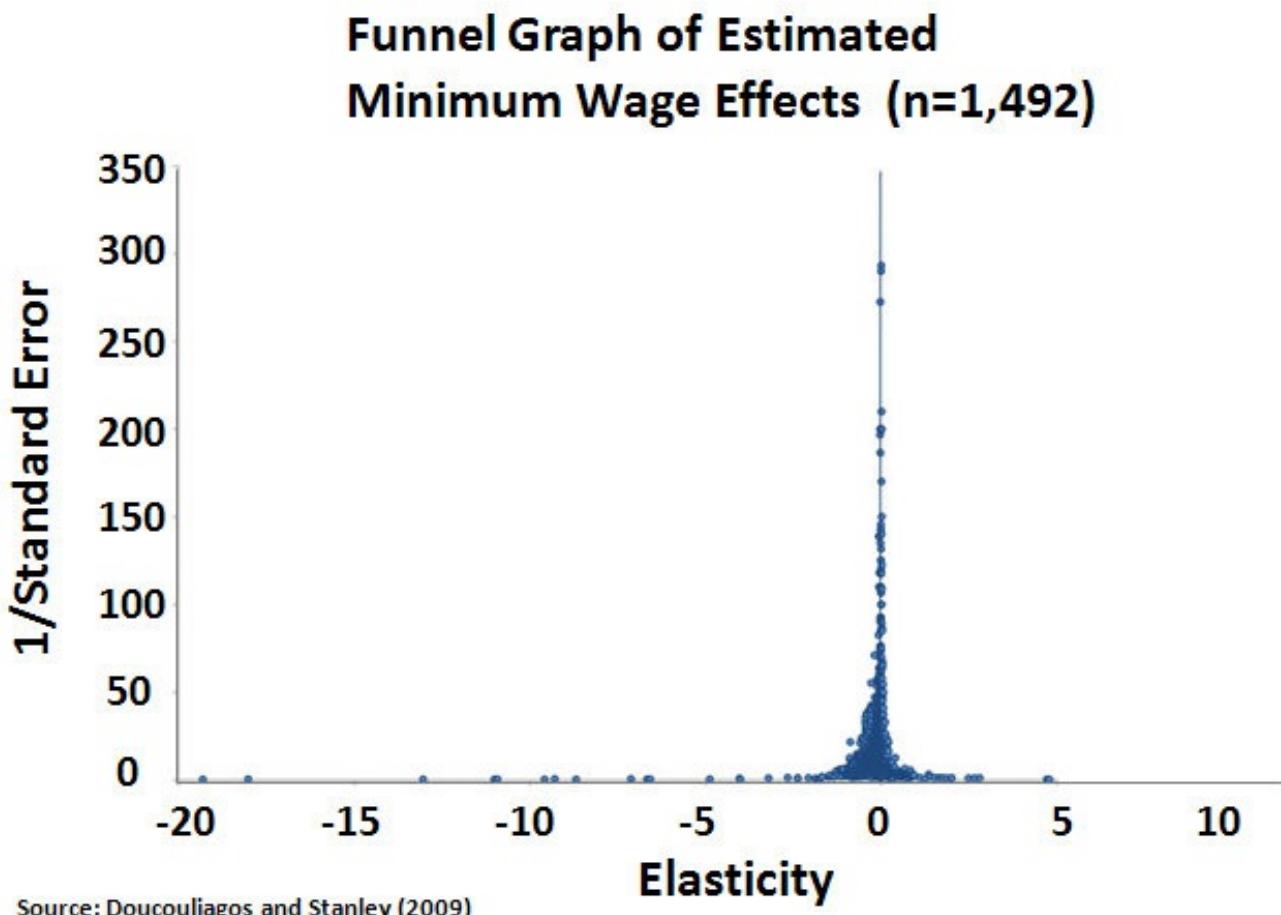
But I worry that most smart people have *not* learned that a list of dozens of studies, several meta-analyses, hundreds of experts, and expert surveys showing almost all academics support your thesis – can *still* be bullshit.

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 interpret 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 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 be 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.

Infinite Debt

Posted on May 10, 2014 by Scott Alexander



I

A patient of mine is getting to that age where she can't support herself independently. She's not a big fan of nursing homes, and I don't blame her. She wants her son to take care of her.

Her son has a career, has a family, doesn't have extra room in his house. Taking care of a sick elderly person is a full-time job, one that can involve everything from giving medications to emptying bedpans. He is not very keen on this plan.

And she says, come on, I worked hard to raise you, you owe me.

And I wonder, how far does this go?

Suppose she is going to need a decade of pretty much twenty-four hour care. She says "Well, I gave you *two* decades of essentially twenty four hour care. You owe me everything, you have to take care of me."

Suppose she lives in a different state from him, and she really doesn't want to leave the family home. She says "Quit your job, sell your house, and move to Michigan to take care of me. You owe me everything, you have to take care of me."

Suppose his wife is really reluctant to share the house with a very demanding sick elderly person. Maybe she knows he works all day and realistically she'll be the one doing the day-to-day caretaking. Maybe she's not so keen on bedpan-emptying. Maybe she categorically refuses. And his mother says "Then divorce your wife and move to Michigan to take care of me. You owe me everything, you have to take care of me."

Most Westerners would probably consider these requests unreasonable. But the mother is probably right. Probably she sacrificed more for her son than she's asking him to sacrifice for her; the credit is still in her favor.

The problem with infinite debts is that they are really hard to repay.

This blog is really bad at staying away from politics for very long, so I'll tell you what this reminds me. This reminds me of the argument some people make against libertarians: "You said you earned your money fair and square. But really, you owe a debt to society. If society hadn't worked really hard inventing things like laws and public health, you would never have been able to found your successful business or even survive past birth. If other members of society hadn't died fighting the Nazis and the Commies and whoever else they fought, you wouldn't have the freedom that helped you succeed. So you owe us, and right now we're calling in that debt in the form of a five percent higher tax rate."

And a lot of libertarians get really angry at that argument because they don't think it's true, they think they don't owe society anything.

I get really angry with that argument for the opposite reason. It is so true that one could presumably owe society *anything*.

What happens if society says “You owe us everything, pay us 10% higher taxes”?

“You owe us everything, pay us 100% higher taxes.”

“You owe us everything, go live in a cardboard box and give us the deed to your house so we can use it as a community center.”

“You owe us everything, I know you want to be a writer but what society really needs right now is oil rig workers, go become an oil rig worker.”

Of course, healthy societies do not say these things – but there have been a lot of unhealthy societies that have. I’m not worried that paying five percent higher taxes is going to lead to me being obligated to quit my job and become an oil rig worker. But the idea of handing society a blank check for anything they want out of me is pretty scary.

I kind of wonder how many of those libertarians who are so upset about a five percent tax raise would be perfectly happy with a Constitutional amendment saying “All rich people must pay 60% of their income in taxes, at which point their entire debt to society is discharged, we promise we will never raise this number above 60% or give them any grief after that, it’s right here in the Constitution.” Even if 60% was way more than a 5% raise over their current rate. 60% may be high, but it is notably lower than “infinity”. Just the acknowledgment that they’re allowed to have their own time and do things they like without being infinitely beholden to society at every moment would be pretty anxiety-relieving.

This is one reason I am so excited about Giving What We Can. Their rule is you give 10% of your income to charity, and you’re allowed in their little club and you get your name on their site as an Officially Recognized Good Person.

For years, I felt like I was *probably* ethically obligated to give all my income to charity, minus whatever I needed to survive. And the fact that I obviously wasn’t going to do that made me not give anything at all.

Once someone told me that my obligation wasn’t infinite, but just some finite amount like ten percent per year, every year, I was thrilled to be able to comply.

And of course there are people who make fun of this. “Oh, you really think you can just give an amount you find “convenient”, then feel like your conscience is clear and you can stop caring and be smug and self-satisfied?”

The proper response to this person is to ask whether they give so much as ten percent.

(“What? No, why should I?! I do *my* part by yelling at you!”)

II

But I think the opposite tendency, the tendency to deny the debt entirely, also falls short of the mark.

I mean, there *are* good arguments for doing so. You never contracted the debt. Your mother never told you as an infant, “I will raise you, but only if you agree to take care of me in any way I require for the rest of my life”, and then made you sign it with your wee little baby hands, threatening to give you back to the stork if you refused. Society

never said “We’ll provide you with public health and technological civilization, but only if you agree to pay any tax rate we set, here’s a ticket to Somalia if you refuse.”

For the decision theoretic take on the question, consider a variation on the Hitchhiker Problem. You’re lying unconscious in the desert, dying of thirst. A very selfish man drives by in his Jeep and considers rescuing you and bringing you to the nearest hospital. This very selfish man is only willing to go through the trouble if you pay him \$100. He decides you probably will pay him \$100 in gratitude for having been rescued, and so takes you to the hospital.

You wake up in the ICU, feeling cool and refreshed. The very selfish man is sitting by your bedside. “Hey,” he says. “I rescued you in the desert because I was pretty sure you’d pay me \$100 for having done so. Will you give me the money?”

I think most of us would feel some obligation to give him the cash. This would be especially true if his actions were a big inconvenience to him – if he had to drive hundreds of miles out of his way, or if he had already paid the \$50 doctor’s bill. Right now a lot of my ideas about morality revolve around “things that help you acausally coordinate hard decision theoretic problems”, and being willing to pay for debts you didn’t contract, as long as you still come out ahead, seems like one of those.

There’s another, more visceral argument. Imagine that my patient didn’t want her son to take 24 hour care of her. She just wanted him to come visit once a year, maybe for Christmas. “Please,” she says, “I’m really lonely and it would mean a lot to me to have you around.”

Now, that guy is under no obligation to go visit random elderly women on Christmas, even if those random elderly women would like company. But most of us would say he *is* under some obligation to visit his mother, or at least that he would be a pretty bad person if he refused. And it’s not just that she is closer to him. Most of us would cash out that obligation in terms of “Your mother did so much for you, can’t you do at least a little for her?”

This is a weird position. You can defend the guy having no debt to his mother. You can defend the guy having an infinite debt to his mother. But a *small* debt to his mother? Where does that come from?

III

I struggle with this concept a lot. I don’t know the moral answer. But I do know the practical answer. Infinite debts make everyone miserable and tend not to be paid at all.

And the moral and practical are sometimes pretty closely aligned. If infinite debts make *everyone* miserable, then by the decision theoretic definition of morality above we might decide to forgive them. After all, every mother was also a child, and it may be that, at the point where you’re making timeless acausal Platonic contracts, everyone agrees to free their child from a debt to them as long as they are themselves freed from debt to their parents. Maybe they would agree that infinite creditors deserve a certain level of respect, very very high respect, but not enough that it ruins your life. This is frustratingly nontechnical. But timeless Platonic contracts are notoriously bad at giving specific figures. Sometimes you’ve just got to seize whatever you can get, find some socially sanctioned middle ground that doesn’t make you feel like you’re a bad person or that you can never enjoy yourself and draw a huge bright line there and defend it to the bitter end.

Maybe your infinite obligation to those worse off than you demands ten percent of your income.

And your infinite obligation to society demands that you pay tax at the prevailing rate.

And your infinite obligation to your mother demands that, at the very least, you call her up on Mother's Day and tell her thanks.

None of those things clear your debt, exactly. But they keep you in good standing. You pay a tiny fraction of your debt, year after year, and it keeps the moral repo man from your door.

The problem with infinite debts is that they are really hard to repay. On the other hand, the *interest* can be quite manageable.

Bottomless Pits Of Suffering

Posted on September 27, 2014 by Scott Alexander



I

A friend on Facebook recently posted the following dilemma, which of course I cannot find right now so I have to vaguely quote my recollection of it:

Would you rather the medieval Church had spent all of its money helping the poor, rather than supporting the arts? So that maybe there were fewer poor people back in medieval times, but we wouldn't have any cathedrals or triptychs or the Sistine Chapel?

I was surprised to see so many people choosing the cathedrals. I think maybe choosing the cathedrals is so appealing because they're right there, you can touch them, but the starving peasants are hidden all the way in the past where you can't see them. So it feels like you're being asked to sacrifice something you really like for something that you would otherwise not have to think about.

This is one of the biggest and scariest problems with utilitarianism. Utilitarianism is at least kind of easy when it's asking you to trade off some things in your normal world for other things in your normal world. But when it asks you to make everything you consider your normal world unambiguously worse to help some other domain you would otherwise never have to think about, then it starts to become unintuitive and scary.

Imagine a happy town full of prosperous people. Every so often they make nice utilitarian decisions like having everyone chip in a few dollars to help someone who's fallen sick, and they feel pretty good about themselves for this.

Then one day an explorer discovers a BOTTOMLESS PIT OF ENDLESS SUFFERING on the outskirts of town. There are hundreds of people trapped inside in a state of abject misery. The Pit Gods agree to release some of their prisoners, but only for appropriately sumptuous sacrifices.

Suddenly the decision isn't just "someone in town makes a small sacrifice to help other people in town". Suddenly it's about the entire town choking off its luxury and prosperity in order to rescue people they don't even know, from this pit they didn't even know was there a week ago. That seems kind of unfair.

So they tell the explorer to cover the lid of the pit with a big tarp that blends in with the surrounding grass, so they don't have to see it, and then go on with their lives.

II

The developing world is *sort of* a bottomless pit of suffering if some First Worlder didn't expect it to be there. But I think most people do expect it to be there, most people are happy to help (a little), and it doesn't really confuse or alarm us too much when we are reminded they still exist and still need help.

But what about nursing homes? Most of the doctors I have talked to agree most nursing homes are terrible. I get a steady trickle of psychiatric patients who are perfectly happy to be in the psychiatric hospital but who *freak out* when I tell them that they seem all better now and it's time to send them back to their nursing home, saying it's terrible and they're abused and neglected and they refuse to go. I very occasionally get elderly patients who have attempted suicide solely because they know doing so will get them out of their nursing home. I don't have a strong feeling for exactly how bad nursing homes are, but everything I have seen is consistent with at least some of them being very bad.

Solving this would be really expensive – I am perpetually surprised at how quietly and effortlessly we seem to soak up nursing home costs that already can run into the tens of thousands of dollars a year. Solving this would also produce no visible gain, in that bedridden old people are very very bad at complaining in ways anyone else can notice, and if we don't want to think about them we don't have to. If we as a country decided to concentrate on decreasing abuse in nursing homes, we might have to take that money away from important causes in our everyday visible world, like welfare and infrastructure and education funding. We would have to take limited Public Attention And Outrage Resources from causes like human rights and gay marriage and what beverages the President is holding while he salutes people. I think everyone agrees it's a lot easier not to think about it, and nobody can make us.

Prisons are an even uglier case. Not only is prison inherently pretty miserable, but there seems to be rampant abuse and violence going on, including at least 5% of prisoners being raped per year. Every couple of weeks there's a new story about how, for example, [prisoners are gouged on phone bills](#) because someone can do it and nobody is stopping them, or how they're kept in cells without air conditioning in 110 degree weather in Arizona because no one has any incentive to change that. If you try to solve it, you'll release prisoners, and at least a few of them are going to commit new crimes which you will have to see and think about. Or you can just whistle, pretend not to notice, and continue to enjoy your nice low-crime society.

III

A lot of the paradoxes of utilitarianism, the things that make it scary and hard to work with, involve philosophers who compulsively seek out bottomless pits and shout at you until you pay attention to them.

Utility monsters are basically one-man bottomless pits.

Pascal's Wager (or Pascal's Mugging, if you prefer) splits the universe into a billion Everett branches, then points out that one of these Everett branches is a bottomless pit and asks the others to make sacrifices to help it.

A lot of the addition paradoxes treat a pool of "potential people" as a bottomless pit.

This seems to be the easiest way to break utilitarianism – point to a bottomless pit, real or imagined, and make everyone in the world lose utility to solve it, forever. It's not always easy to come up with solutions that successfully rule out these problems, while preserving our intuition that we should continue to worry about people in nursing homes or jails.

Contractualism scares me a little because it offers *too easy* an out from bottomless-pit type dilemmas. It seems really easy to say "All of us people not in jail, we'll agree to look out for one another, and as for those guys, screw them". You would need to have something like a veil of ignorance, or at least a [good simulation of one](#), to even begin to care.

Nydwracu's Fnords

Posted on May 24, 2014 by Scott Alexander



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](#)

The Virtue of Silence

Posted on June 14, 2013 by Scott Alexander



Leah Libresco writes a couple of essays ([1](#), [2](#)) on an ethical dilemma reported in the New York Times. In the course of a confidential medical history, a doctor hears her patient is suffering from stress-related complaints after having sent an innocent man to prison. The doctor wants to know whether it is ethical to report the matter to the police. The Times' columnist says yes – it would save the poor prisoner. Leah says no – violating medical confidentiality creates an expectation that medical confidentiality will be violated in the future, thus dooming patients who are too afraid to talk about drug use or gay sex or other potentially embarrassing but important medical risk factors.

But both sides are ignoring the much bigger dilemma lurking one meta-level up: is it ethical to debate this dilemma in the *New York Times*?

Let's look more closely at that phrase "violating medical confidentiality creates an expectation that medical confidentiality will be violated in the future." There's a very abstruse angels-and-clockwork interpretation of "creates an expectation" where, by making the decision to violate confidentiality, you are altering the Platonic machinery of the Universe in a way that allows other beings who know your source code to determine that you will do this. But most people don't have the [decision theory](#) to understand this, and anyway most doctors do not publish their source code online.

The way "creates an expectation" pans out in *our* universe is that somebody hears that a doctor violated medical confidentiality, and that person tells someone else, and that person tells someone else, until eventually someone who was going to tell their doctor about having gay sex with drugs remembers having heard the story and decides not to.

How exactly would people hear about this doctor who revealed the innocence of the prisoner? Through the ensuing court case? Nah. Most people wouldn't obsessively read the minutes of every single case at the local courthouse [unless of course it has something to do with gender](#). Really, the only way that someone could hear about a doctor violating medical confidentiality is if she, like, somehow got a description of her intention to do so published in meticulous detail in the *New York Times*.

Oh, *right*.

The entire negative effect of the doctor breaking her promise is that it would make people doubt medical confidentiality in the future. But *whether or not the doctor ends up breaking her promise*, thousands of New York Times readers now know that doctors strongly consider breaking medical confidentiality, and that ethics columnists tell them it's okay to do so. It seems like the whether the doctor actually keeps her promise or not in this particular case is of minuscule importance compared to the damage that the column has already done.

Silence is a *hard* virtue. All the other virtues have the advantage that, when you practice them, people will praise you. Sometimes if your moral system is very different from your friends' people will attack you for your virtues, but [getting attacked by sufficiently horrible people](#) can sometimes be just as gratifying as praise. But if you stay silent, there's no praise *and* no attacks. By definition, no one even knows you made a courageous moral choice.

(Eliezer mentioned in the comments of my [Against Bravery Debates](#) that he's spent a couple decades pushing ideas almost everyone else thinks are crackpot, and he's never appealed to bravery at all. He is one hundred percent correct

and I have one hundred percent never noticed despite reading almost everything he's written for several years. That's the Virtue of Silence for you.)

(I had like five much better examples here, all of which would be very clever, and each time I had to catch myself and say "Wait a second, by bringing that up I'm violating the virtue I'm supposed to be pushing here, aren't I?")

One example of silence I deeply appreciate is *people who don't talk about the latest viral issue*. I'm trying to think of an example that's not too destructive to bring up... hmmmm... go for something old... [Elevatorgate!](#) Nearly everyone who talked about Elevatorgate mentioned that it was outrageous that the blogosphere was making such a big deal about it, missing the similarity to the old adage that "you aren't stuck in traffic, you *are* traffic." Somewhere there was someone who wanted to write about Elevatorgate, thought about it, and *decided not to*. That person deserves the sincere thanks of a grateful Internet.

So having made the case for the other side of the confidentiality-newspaper meta-dilemma, am I *actually* pushing the claim that it is a moral law not to publicize information that could have bad consequences?

But I notice that this sort of thing almost always ends up making people angry and having a perverse effect where demands not to draw Mohammed turn into Everyone Draw Mohammed Day (see: [Streisand Effect](#)). It also sometimes snowballs to the point where not only can you not talk about X, but you can't talk about the demand not to talk about X because that would be referring to X obliquely, and you can't talk about the demand not to talk about the demand to talk about X, until eventually you climb up so many meta-levels that you collapse from hypoxia and have to be rescued by Sherpas. Then you get a "callout culture" where people try to gain easy Virtue Points by telling people discussing issues that they Should Not Be Discussing Them and other people try to gain easy Virtue Points by being the Brave Defender of Freedom of Speech.

And maybe that's useful if it's something like gender where everyone wants to talk about it all the time anyway, but we don't really need to do that to medical confidentiality, do we?

Maybe this is one of those rare cases where the word "supererogatory" might be useful. Yelling at people who talk about violations of medical confidentiality would just lead to "ARE OUR DISCUSSIONS OF MEDICAL CONFIDENTIALITY BEING SILENCED??!?" on the front page of the *New York Times*. And fretting over talking about it with your friends, or publishing a blog article about it (cough) is probably on the moral level of those Jains who walk everywhere with a broom in front of them so that they don't accidentally squash any bugs. But if someone is really really concerned about it and wants to be a great person, then yeah, I think writing to the *New York Times* about it requires a bit of thought.

And since I *am* publishing a blog article about it (VIRTUE OF SILENCE IS REALLY HARD!) let me restore some Virtue Points by confirming that I will *not* betray private patient information of this sort if such a dilemma comes up except when legally required. Trust me, I'm a doctor.

Fundamental Value Differences Are Not That Fundamental

Posted on July 18, 2018 by Scott Alexander



I

[Ozy](#) (and [others](#)) talk about fundamental value differences as a barrier to cooperation.

On their model (as I understand it) there are at least two kinds of disagreement. In the first, people share values but disagree about facts. For example, you and I may both want to help the Third World. But you believe foreign aid helps the Third World, and I believe it props up corrupt governments and discourages economic self-sufficiency. We should remain allies while investigating the true effect of foreign aid, after which our disagreement will disappear.

In the second, you and I have fundamentally different values. Perhaps you want to help the Third World, but I believe that a country should only look after its own citizens. In this case there's nothing to be done. You consider me a heartless monster who wants foreigners to starve, and I consider you a heartless monster who wants to steal from my neighbors to support random people halfway across the world. While we can agree not to have a civil war for pragmatic reasons, we shouldn't mince words and pretend not to be enemies. Ozy writes (liberally edited, read the original):

From a conservative perspective, I am an incomprehensible moral mutant... however, from my perspective, conservatives are perfectly willing to sacrifice things that actually matter in the world – justice, equality, happiness, an end to suffering – in order to suck up to unjust authority or help the wealthy and undeserving or keep people from having sex lives they think are gross.

There is, I feel, opportunity for compromise. An outright war would be unpleasant for everyone... And yet, fundamentally... it's not true that conservatives as a group are working for the same goals as I am but simply have different ideas of how to pursue it... my read of the psychological evidence is that, from my value system, about half the country is evil and it is in my self-interest to shame the expression of their values, indoctrinate their children, and work for a future where their values are no longer represented on this Earth. So it goes.

And from the subreddit comment by GCUPokeItWithASStick:

I do think that at a minimum, if you believe that one person's interests are intrinsically more important than another's (or as the more sophisticated versions play out, that ethics is agent-relative), then something has gone fundamentally wrong, and this, I think, is the core of the distinction between left and right. Being a rightist in this sense is totally indefensible, and a sign that yes, you should give up on attempting to ascertain any sort of moral truth, because you can't do it.

I will give this position its due: I agree with the fact/value distinction. I agree it's conceptually very clear what we're doing when we try to convince someone with our same values of a factual truth, and confusing and maybe impossible

to change someone's values.

But I think the arguments above are overly simplistic. I think rationalists might be especially susceptible to this kind of thing, because we often use economic models where an agent (or AI) has a given value function (eg "produce paperclips") which generates its actions. This kind of agent really does lack common ground with another agent whose goal function is different. But humans rarely work like this. And even when they do, it's rarely in the ways we think. We are far too quick to imagine binary value differences that line up exactly between Us and Them, and far too slow to recognize the complicated and many-scaled pattern of value differences all around us.

Eliezer Yudkowsky writes, in [Are Your Enemies Innately Evil?](#):

On September 11th, 2001, nineteen Muslim males hijacked four jet airliners in a deliberately suicidal effort to hurt the United States of America. Now why do you suppose they might have done that? Because they saw the USA as a beacon of freedom to the world, but were born with a mutant disposition that made them hate freedom?

Realistically, most people don't construct their life stories with themselves as the villains. Everyone is the hero of their own story. The Enemy's story, as seen by the Enemy, is not going to make the Enemy look bad. If you try to construe motivations that would make the Enemy look bad, you'll end up flat wrong about what actually goes on in the Enemy's mind.

So what was going through the 9/11 hijackers' minds? How many value differences did they have from us?

It seems totally possible that the hijackers had no value differences from me at all. If I believed in the literal truth of Wahhabi Islam – a factual belief – I might be pretty worried about the sinful atheist West. If I believed that the West's sinful ways were destroying my religion, and that my religion encoded a uniquely socially beneficial way of life – both factual beliefs – I might want to stop it. And if I believed that a sufficiently spectacular terrorist attack would cause people all around the world to rise up and throw off the shackles of Western oppression – another factual belief – I might be prepared to sacrifice myself for the greater good. If I thought complicated Platonic contracts of cooperation and nonviolence didn't work – sort of a factual belief – then my morals would no longer restrain me.

But of course maybe the hijackers had a bunch of value differences. Maybe they believed that American lives are worth nothing. Maybe they believed that striking a blow for their homeland is a terminal good, whether or not their homeland is any good or its religion is true. Maybe they believe any act you do in the name of God is automatically okay.

I have no idea how many of these are true. But I would hate to jump to conclusions, and infer from the fact that they crashed two planes that they believe crashing planes is a terminal good. Or infer from someone opposing abortion that they [just think oppressing women is a terminal value](#). Or infer from people committing murder that they [believe in murderism](#), the philosophy that says that murder is good. I think most people err on the side of being too quick to dismiss others as fundamentally different, and that a little charity in assessing their motives can go a long way.

II

But that's too easy. What about people who didn't die in self-inflicted plane crashes, and who can just tell us their values? Consider the original example – foreign aid. I've heard many isolationists say in no uncertain terms that they believe we should not spend money to foreign countries, and that this is a basic principle and not just a consequence of some factual belief like that foreign countries would waste it. Meanwhile, I know other people who argue that we

should treat foreigners exactly the same as our fellow citizens – indeed, that it would be an affront to basic compassion and to the unity of the human race not to do so. Surely this is a strong case for actual value differences?

My only counter to this line of argument is that almost nobody, me included, ever takes it seriously or to its logical conclusion. I have never heard any cosmopolitans seriously endorse the idea that the Medicaid budget should be mostly redirected from the American poor (who are already plenty healthy by world standards) and used to fund clinics in Africa, where a dollar goes much further. Perhaps this is just political expediency, and some would talk more about such a plan if they thought it could pass. But in that case, they should realize that they are very few in number, and that their value difference isn't just with conservatives but with the overwhelming majority of their friends and their own side.

And if nativist conservatives are laughing right now, I know that some of them have given money to foreign countries affected by natural disasters. Some have even suggested the government do so – when the US government sent resources to Japan to help rescue survivors of the devastating Fukushima tsunami, I didn't hear anyone talk about how those dollars could better be used at home.

Very few people have consistent values on questions like these. That's because nobody naturally has principles. People take the unprincipled mishmash of their real opinions, extract principles out of it, and follow those principles. But the average person only does this very weakly, to the point of having principles like "it's bad when you lie to me, so maybe lying is wrong in general" – and even moral philosophers do it less than a hundred percent and apply their principles inconsistently.

(this isn't to say those who have consistent principles are necessarily any better grounded. I've talked a lot about shifting views of federalism: when the national government was against gay marriage, conservatives supported top-down decision-making at the federal level, and liberals protested for states' rights. Then when the national government came out in support, conservatives switched to wanting states' rights and liberals switched to wanting top-down federal decisions. We can imagine some principled liberal who, in 1995, said "It seems to me right now that state rights are good, so I will support them forevermore, even when it hurts my side". But her belief still would have ended up basically determined by random happenstance; in a world where the government started out supporting gay marriage but switched to oppose it, she would have – and stick to – the opposite principle)

But I'm saying that what principle you verbalize ("I believe we must treat foreigners exactly as our own citizens!") isn't actually that interesting. In reality, there's a wide spectrum of what people will do with foreigners. If we imagine it as a bell curve, the far right end has a tiny number of hyper-consistent people who oppose any government money going abroad unless it directly helps domestic citizens. A little further towards the center we get the people who say they believe this, but will support heroic efforts to rescue Japanese civilians from a tsunami. The bulge in the middle is people who want something like the current level of foreign aid, as long as it goes to sufficiently photogenic children. Further to the left, we get the people I'm having this discussion with, who usually support something like a bit more aid and open borders. And on the far left, we get another handful of hyper-consistent people, who think the US government should redirect the Medicaid budget to Africa.

If you're at Point N in some bell curve, how far do you have to go before you come to someone with "fundamental value differences" from you? How far do you have to go before someone is inherently your enemy, cannot be debated with, and must be crushed in some kind of fight? If the answer is "any difference at all", I regret to inform you that the bell curve is continuous; there may not be anyone with exactly the same position as you.

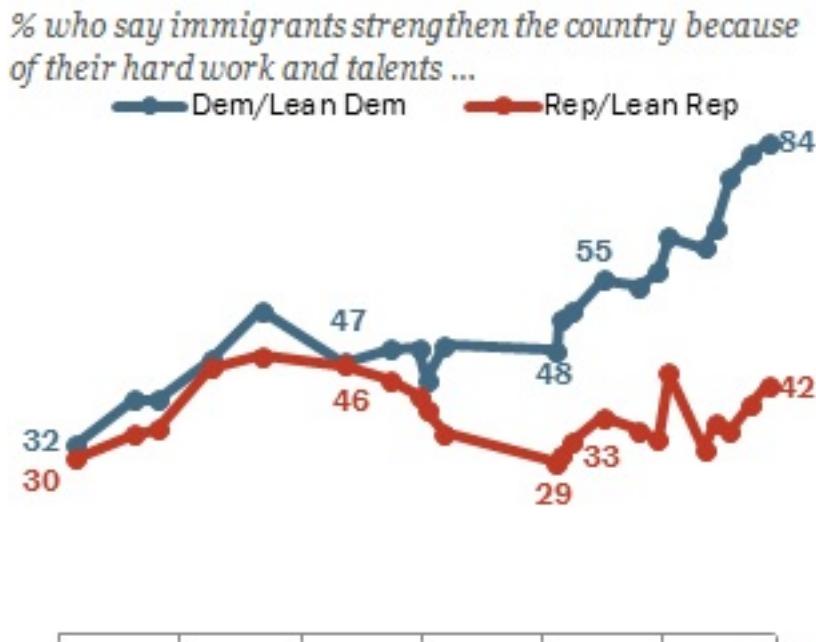
And that's just the one issue of foreign aid. Imagine a hundred or a thousand such issues, all equally fraught. God help GCU, who goes further and says you're "indefensible" if you believe *any* human's interests are more important than any other's. Does he (I'll assume it's a he) do more to help his wife when she's sick than he would to help a random stranger? This isn't meant to be a gotcha, it's meant to be an example of how we formulate our morality. Person A

cares more about his wife than a random person, and also donates some token amount to help the poor in Africa. He dismisses caring about his wife as noise, then extrapolates from the Africa donation to say “we must help all people equally”. Person B also cares more about his wife than a random person, and also donates some token amount to Africa. He dismisses the Africa donation as noise, then extrapolates from his wife to “we must care most about those closest to us”. I’m not saying that how each person frames his moral principle won’t have effects later down the line, but those effects will be the tail wagging the dog. If A and B look at each other and say “I am an everyone-equally-er, you are a people-close-to-you-first-er, we can never truly understand one another, we must be sworn enemies”, they’re putting a whole lot more emphasis on which string of syllables they use to describe their mental processes than really seems warranted.

Why am I making such a big deal of this? Isn’t a gradual continuous value difference still a value difference?

Yes. But I expect that (contra the Moral Foundations idea) both the supposed-nativist and the supposed-cosmopolitan have at least a tiny bit of the instinct toward nativism and the instinct toward cosmopolitanism. They may be suppressing one or the other in order to fit their principles. The nativist might be afraid that if he admitted any instinct toward cosmopolitanism, people could force him to stop volunteering at his community center, because his neighbor’s children are less important than starving Ethiopians and he should be helping them somehow instead. The cosmopolitan might be afraid that if he admitted any instinct toward preferring people close to him, it would justify a jingoistic I’ve-got-mine attitude that thinks of foreigners as subhuman.

But the idea that they’re inherently different, and neither can understand the other’s appeals or debate each other, is balderdash. A lot of the our-values-are-just-inherently-different talk I’ve heard centers around immigration. Surely liberals must have some sort of strong commitment to the inherent moral value of foreigners if they’re so interested in letting them into the country? Surely conservatives must have some sort of innate natives-first mentality to think they can just lock people out? But...



Source: Survey conducted June 8-18 and June 27-July 9, 2017.
Q25g

PEW RESEARCH CENTER

Okay. I admit this is a question about hard work and talents, which is a factual question. But we both know that you would get basically the same results if you asked "IMMIGRATION GOOD OR BAD?" or "DO IMMIGRANTS HAVE THE SAME RIGHTS TO BE IN THIS COUNTRY AS THE NATIVE BORN?" or whatever. And what we see is that this is totally contingent and dependent on the politics of the moment. Of all those liberals talking about how they can't possibly comprehend conservatives because being against immigration would just require completely alien values, half of them were anti-immigrant ten years ago. Of all those conservatives talking about how liberals can never be convinced by mere debate because debate can't cut across fundamental differences, they should try to figure out why their own party was half again as immigrant-friendly in 2002 as in 2010.

I don't think anyone switched because of anything they learned in a philosophy class. They switched because it became mildly convenient to switch, and they had a bunch of pro-immigrant instincts and anti-immigrant instincts the whole time, so it was easy to switch which words came out of their mouths as soon as it became convenient to do so.

So if the 9/11 hijackers told me they truly placed zero value on American lives, I would at least reserve the possibility that sure, this is something you say when you want to impress your terrorist friends, but that in a crunch – if they saw an anvil about to drop on an American kid and had only a second to push him out of the way – they would end up having some of the same instincts as the rest of us.

III

Is there anyone at all whom I am willing to admit definitely, 100%, in the most real possible way, has different values than I do?

I think so. I remember a debate I had with my ex-girlfriend. Both of us are atheist materialist-computationalist utilitarian rationalist effective altruist liberal-tarians with 99% similar views on every political and social question. On the other hand, it seemed axiomatic to me that it wasn't morally good/obligatory to create extra happy people (eg have a duty to increase the population from 10,000 to 100,000 people in a way that might eventually create the [Repugnant Conclusion](#)), and it seemed equally axiomatic to her that it *was* morally good/obligatory to do that. We debated this maybe a dozen times throughout our relationship, and although we probably came to understand each other's position a little more, and came to agree it was a hard problem with some intuitions on both sides, we didn't come an inch closer to agreement.

I've had a few other conversations that ended with me feeling the same way. I may not be the typical Sierra Club member, but I consider myself an environmentalist in the sense of liking the environment and wanting it to be preserved. But I don't think I value biodiversity for its own sake – if you offered me something useful in exchange for half of all species going extinct – promising that they would all be random snails, or sponges, or some squirrel species that looked exactly like other squirrel species, or otherwise not anything we cared about – I'd take it. If you offered me all charismatic megafauna being relegated to zoos in exchange for lots of well-preserved beautiful forests that people could enjoy whenever they wanted, I would take that one too. I know other people who consider themselves environmentalists who are horrified by this. Some of them agree with me on every single political issue that real people actually debate.

I think these kinds of things are probably real fundamental value difference. But if I'm not sure I have any fundamental value differences with the 9-11 hijackers, and I am sure I have one with one of the people I'm closest to in the entire world, how big a deal is it, exactly? The world isn't made of Our Tribe with our fundamental values and That Tribe There with their fundamental values. It's made of a giant mishmash of provisional things that solidify into values at some point but can be unsolidified by random chance or temporary advantage, and everyone probably has a couple unexplored value differences and unexpected value similarities with everyone else.

This means that trying to use shaming and indoctrination to settle value differences is going to be harder than you think. Successfully defeat the people on the other side of the One Great Binary Value Divide That Separates Us Into Two Clear Groups, and you're going to notice you still have some value differences with your allies (if you don't now, you will in ten years, when the political calculus changes slightly and their deepest ethical beliefs become totally different). Beat your allies, and you and the subset of remaining allies will *still* have value differences. It's value differences all the way down. You will have an infinite number of fights, and you're sure to lose some of them. Have you considered [getting principles](#) and [using asymmetric weapons?](#)

I'm not saying you don't have to fight for your values. The foreign aid budget still has to be some specific number, and if your explicitly-endorsed principles disagree with someone else's explicitly-endorsed principles, then you've got to fight them to determine what it is.

But "remember, liberals and conservatives have fundamental value differences, so they are two tribes that can't coexist" is the wrong message. "Remember, everyone has weak and malleable value differences with everyone else, and maybe a few more fundamental ones though it's hard to tell, and neither type necessarily line up with tribes at all, so they had damn well better learn to coexist" is more like it.

Beware Isolated Demands For Rigor

August 14, 2014



I

From *Identity, Personal Identity, and the Self* by John Perry:

There is something about practical things that knocks us off our philosophical high horses. Perhaps Heraclitus really thought he couldn't step in the same river twice. Perhaps he even received tenure for that contribution to philosophy. But suppose some other ancient had claimed to have as much right as Heraclitus did to an ox Heraclitus had bought, on the grounds that since the animal had changed, it wasn't the same one he had bought and so was up for grabs. Heraclitus would have quickly come up with some ersatz, watered-down version of identity of practical value for dealing with property rights, oxen, lyres, vineyards, and the like. And then he might have wondered if that watered-down vulgar sense of identity might be a considerably more valuable concept than a pure and philosophical sort of identity that nothing has.

Okay, but I can think of something worse than that.

Imagine Heraclitus as a cattle rustler in the Old West. Every time a rancher catches him at his nefarious business, he patiently explains to them that identity doesn't exist, and therefore the same argument against private property as made above. Flummoxed, they're unable to think of a response before he rides off into the sunset.

But then when Heraclitus *himself* needs the concept of stable personal identity for something – maybe he wants to deposit his ill-gotten gains in the bank with certainty that the banker will give it back to him next time he shows up to withdraw it, or maybe he wants to bribe the sheriff to ignore his activities for the next while – all of a sudden Heraclitus is willing to tolerate the watered-down vulgar sense of identity like everyone else.

(actually, I can think of something even worse than *that*, which is a TV western based on this premise, where a roving band of pre-Socratic desperadoes terrorizes Texas. The climax is no doubt when the hero strides onto Main Street, revolver in hand, saying "There's a new sheriff in town." And Parmenides gruffly responds "No, I'm pretty sure that's impossible.")

At its best, philosophy is a revolutionary pursuit that dissolves our common-sense intuitions and exposes the possibility of much deeper structures behind them. One can respond by becoming a saint or madman, or by becoming a pragmatist who is willing to continue to participate in human society while also understanding its theoretical limitations. Both are respectable career paths.

The problem is when someone chooses to apply philosophical rigor *selectively*.

Heraclitus could drown in his deeper understanding of personal identity and become a holy madman, eschewing material things and taking no care for the morrow because he does not believe there is any consistent self to experience it. Or he could engage with it from afar, becoming a wise scholar who participating in earthly affairs while

drawing equanimity from the realization that there is a sense in which all his accomplishments will be impermanent.

But if he only applies his new theory when he wants other people's cows, *then* we have a problem. Philosophical rigor, usually a virtue, has been debased to an *isolated demand for rigor* in cases where it benefits Heraclitus.

A fair use of philosophical rigor would prevent both Heraclitus and his victims from owning property, and thus either collapse under its own impracticality or usher in a revolutionary new form of economic thinking. An isolated demand for philosophical rigor, applied by Heraclitus to other people but never the other way around, would merely give Heraclitus an unfair advantage in the existing system.

II

A while ago I wrote a post called [Military Strikes Are An Extremely Cheap Way To Help Foreigners](#) which was a response to a Matt Yglesias post called [the opposite](#). Yglesias was opposed to "humanitarian" military intervention (think the air strikes on ISIS going on right now, justified under the cause of preventing a genocide) and his argument was that this was extremely cost-ineffective compared to just giving the money to GiveWell's top-rated charity – at the time he was writing, malaria prevention.

I argued he was wrong about his numbers. But I also argued he was unfairly making an isolated demand for philosophical rigor.

Once you learn about utilitarianism and effective charity, you can become the holy madman, donating every cent you have beyond what is strictly necessary to survive and hold down a job to whatever the top rated charity is.

Or you can become the worldly scholar, continuing to fritter away your money on things like "hot water" and "food other than gruel" but appreciating the effective-utilitarian perspective and trying to make a few particularly important concessions to it.

Or you can use it to steal other people's cows. This is what I accused Matt Yglesias of doing. Presumably there are lots of government programs Yglesias supports – I suggested PBS – and he would never *dream* of demanding that we defund them in the hopes of donating the money to malaria prevention. But if for political reasons he doesn't support air strikes, suddenly *that* plan has to justify itself according to rigorous criteria that no government program that exists could possibly pass.

Government spending seems to be a particularly fertile case for this problem. I remember hearing some conservatives complain: sex education in public schools is an outrage, because *my* tax dollars are going to support something I believe is morally wrong.

This is, I guess, a demand for ethical rigor. That no one should ever be forced to pay for something they don't like. Apply it consistently, and conservatives shouldn't have to pay for sex ed, liberals shouldn't have to pay for wars, and libertarians shouldn't have to pay for *anything*, except maybe a \$9.99 tax bill yearly to support the police and a minimal court system.

Applied consistently, you become the holy madman demanding either total anarchy or some kind of weird system of tax earmarks which would actually be pretty fun to think about. Or the worldly scholar with a strong appreciation for libertarian ideas who needs a really strong foundational justification for spending government money on things that a lot of people oppose.

Applied *inconsistently*, you're just stealing cows again, coming up with a clever argument against the programs you

don't like while defending the ones you do.

III

But this is the sort of uncouth behavior we expect of political partisans. What about *science*?

Suppose there are scientists on both sides of a controversial issue – for example, economists studying the minimum wage. One team that supports a minimum wage comes up with a pretty good study showing with $p < 0.05$ that minimum wages help the economy in some relevant way.

The Science Czar (of course we have a science czar! We're not monsters!) notes that $p < 0.05$ is really a shoddy criterion that can prove anything and they should come back when they have $p < 0.01$. I have a *huge* amount of sympathy with the Science Czar on this one, by the way.

Soooo the team of economists spends another five years doing another study and finds with $p < 0.01$ that the minimum wage helps the economy in some important way.

The Science Czar notes that their study was correlational only, and that correlational studies suck. We really can't show that minimum wages are any good without a randomized controlled trial.

Luckily, the governments of every country in the world are totally game for splitting their countries in half and instituting different economic regimes in each part for ten years, so after a decade it comes out that in the randomized controlled trial the minimum wage helped the economy with $p < 0.01$.

The Science Czar worries about publication bias. What if there were a lot of other teams who got all the countries in the world to split in half and institute different wage policies in each of the two territories for one decade, but they weren't published because their results weren't interesting enough?

Everything the Science Czar has said so far makes perfect sense and he is to be commended for his rigor and commitment to the job. Science is really hard and even tiny methodological mistakes can in principle invalidate an entire field.

But now suppose that a team shows that, in a sample of six restaurants in Podunk Ohio, there was a nonsignificant trend towards the minimum wage making things a little worse.

And the Science Czar says: awesome! That solves that debate, minimum wage is bad, let's move on to investigating nominal GDP targeting.

Now it looks like the Science Czar is just a jerk who's really against minimum wage. All his knowledge of the standards of scientific rigor are going not towards bettering science, but toward worsering science. He's not trying to create a revolutionary new scientific regime, he's taking pot shots.

I see this a lot in medicine. Someone jumps on a new study showing the selenium or chromium or plutonium or whatever cures cancer. It is brought up that no, really, the medical community has investigated this sort of thing before, and it has always been found that it doesn't.

"Well, maybe the medical community wasn't investigating it the right way! Maybe the investigators were biased! Maybe they didn't randomize right! Maybe they used a population unusually susceptible to cancer-getting! Ninety percent of medical studies are wrong! Those twenty experiments showing a lack of effect could be total bunk!"

Yes, maybe these things happened in each of the twenty studies that disagree with you.

Or maybe they happened in the one contrarian study you are getting so excited about.

IV

The unholy combination of isolated demands for philosophical rigor and isolated demands for scientific rigor is isolated demands for mathematical-statistical-conceptual rigor, ie the sort of thing this blog has been talking about all week.

I have already [been made fun of](#) for how many different things I am metaphorically comparing IQ to – speed, blood pressure, [comas](#) – so I guess it can't hurt to add another example I only thought of today. How about crime? It's usually measured by crime rate – a made-up statistic that combines subfactors like arson (maybe higher when fire insurance pays out better), property damage (maybe higher during periods of ethnic tension and frequent riots) and theft (maybe higher when income inequality is worse). There is assumed to be a General Factor Of Crime (presumably caused by things like poor policing, dark alleys, broken families, et cetera) but I would be extremely surprised if anyone had ever proven Beyond A Shadow Of A Doubt that the factor analysis works out here.

When Cosma Shalizi says [he's not sure](#) about the factor analysis in IQ, I have no quarrel with him, because Cosma Shalizi's response to *everything in the world* is to glare at it for not being sufficiently statistically rigorous.

But when other people are totally happy to talk about speed and blood pressure and comas and the crime rate, and then suddenly switch to a position that we can't talk about IQ *at all* unless we have a perfect factor-analytical proof of its obeying certain statistical rules, then I worry they're just out to steal cows.

Likewise, if someone were to just never acknowledge any sorts of groups of objects except those that could be statistically proven to fall out into absolutely separate clusters in which variance within each cluster is less than variance between clusters, well, at least they would be fun to talk to at dinner parties.

But when people never even begin to question the idea of different cultures but make exacting demands of anyone before they can talk about different races – even though [the two ideas are statistically isomorphic](#) – then I think they're just out to steal cows.

So this is another technique for avoiding [Eulering](#) – is your interlocutor equally willing to apply their complex mathematical argument to everything else.

I think if I hadn't known anything about Bayesian probability, I would have examined the McGrews' Bayesian argument for the Gospels by seeing if it applied equally well to [Mormonism](#), the [control group for Christianity](#).

V

The old man stamped his boot in the red dirt, kicking up a tiny cloud of dust. "There's a new sheriff in town," he told them.

"No, I'm pretty sure that's impossible," says Parmenides. "There's no such thing as change, only the appearance thereof."

"Well then," says the old man, "I reckon you won't mind the false illusion of your surroundings appearing to change into a jail cell." And he took out his six-shooter and held it steady.

"Hold on," said Thales. "We don't want any trouble here. All is water, so all we did was steal a little bit of water from people. We can give you some water back, and everything will be even, right?" He gestured to a watering trough for horses on the side of the street, which was full of the stuff.

"Just so long as you don't mind being sprayed with some very hard water from my squirt gun," the old man answered, and the six-shooter was pointed at the Milesian now.

"Ha!" said Zeno of Elea. "You don't scare us. In order to hit Thales, your bullet would have to get halfway to him, then half of the remaining distance, and so on. But that would require an infinite number of steps, therefore it is impossible."

"Sorry," said the old man, "I couldn't hear you because it's logically impossible for the sound waves encoding your speech to reach my ears."

"We're not even the same people as the guys who stole those cattle!" said Heraclitus. "Personal identity is an illusion!"

"Then you won't mind coming to the courthouse with me," replied the old man "to help the judge imprison some other people who look just like you."

The last of them, the tall one, said nothing. He just raised his revolver in a fluid motion and shot at the old man.

The old man saw it coming and jumped out of the way. The air was briefly full of bullets. Bang! Thales went down! Bang bang! Heraclitus! Bang bang! Parmenides and Zeno. Bang bang bang! The old man was hit in the arm, but still standing. Bang bang bang bang...

It was just the old man and the tall one now. The tall one picked up his gun and fired. Nothing happened. Out of bullets.

The old man smiled wryly, his six-shooter still in his hand.

"I know what you're thinking. You're thinking – did he fire six shots, or only five? Well, you've got to ask yourself a question – do you feel lucky? Well, do you, punk?"

The tall one didn't budge. "Man is the measure of all things," said Protagoras. "If I believe you fired six shots, then by my personal epistemic standards, you fired six shots."

The old man didn't say anything.

"You see," the Sophist continued. "Out of all of them, I alone was truly consistent. They all came up with clever theories, then abandoned them whenever it conflicted with their self-interest. I was more honest. I just said at the beginning that my self-interest determined truth, and so never suffered any temptation to depart from my position."

The old man took off the bandana covering his face. "Man may be the measure of all things. But I've taken *your* measure, Protagoras, and found it wanting."

"Socrates?!" the Sophist gasped.

"The only truly consistent people are the dead, Protagoras," he said – and squeezed the trigger.

Society Is Fixed, Biology Is Mutable

Posted on September 10, 2014 by Scott Alexander



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 responsibility 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.

The Parable Of The Talents

Posted on January 31, 2015 by Scott Alexander



Content note: scrupulosity and self-esteem triggers, IQ, brief discussion of weight and dieting. Not good for growth mindset.

I

I sometimes blog about research into IQ and human intelligence. I think most readers of this blog already know [IQ is 50% to 80% heritable](#), and that it's so important for intellectual pursuits that [eminent scientists in some fields have average IQs around 150 to 160](#). Since IQ this high only appears in 1/10,000 people or so, it beggars coincidence to believe this represents anything but a very strong filter for IQ (or something correlated with it) in reaching that level. If you saw a group of dozens of people who were 7'0 tall on average, you'd assume it was a basketball team or some other group selected for height, not a bunch of botanists who were all very tall by coincidence.

A lot of people find this pretty depressing. Some worry that taking it seriously might damage the "growth mindset" people need to fully actualize their potential. This is important and I want to discuss it eventually, but not now. What I want to discuss now is people who feel *personally* depressed. For example, a comment from last week:

I'm sorry to leave self a self absorbed comment, but reading this really upset me and I just need to get this off my chest... How is a person supposed to stay sane in a culture that prizes intelligence above everything else – especially if, as Scott suggests, Human Intelligence Really Is the Key to the Future – when they themselves are not particularly intelligent and, apparently, have no potential to ever become intelligent? Right now I basically feel like pond scum.

I hear these kinds of responses every so often, so I should probably learn to expect them. I never do. They seem to me precisely backwards. There's a moral gulf here, and I want to throw stories and intuitions at it until enough of them pile up at the bottom to make a passable bridge. But first, a comparison:

Some people think body weight is biologically/genetically determined. Other people think it's based purely on willpower – how strictly you diet, how much you can bring yourself to exercise. These people get into some pretty acrimonious debates.

Overweight people, and especially people who feel unfairly stigmatized for being overweight, tend to cluster on the biologically determined side. And although not all believers in complete voluntary control of weight are mean to fat people, the people who are mean to fat people pretty much all insist that weight is voluntary and easily changeable.

Although there's a lot of debate over the science here, there seems to be broad agreement on both sides that the more compassionate, sympathetic, progressive position, the position promoted by the kind of people who are really worried about stigma and self-esteem, is that weight is biologically determined.

And the same is true of mental illness. Sometimes I see depressed patients whose families *really* don't get it. They say "Sure, my daughter feels down, but she needs to realize that's no excuse for shirking her responsibilities. She needs to

just pick herself up and get on with her life.” On the other hand, most depressed people say that their depression is more fundamental than that, not a thing that can be overcome by willpower, certainly not a thing you can just ‘shake off’.

Once again, the compassionate/sympathetic/progressive side of the debate is that depression is something like biological, and cannot easily be overcome with willpower and hard work.

One more example of this pattern. There are frequent political debates in which conservatives (or straw conservatives) argue that financial success is the result of hard work, so poor people are just too lazy to get out of poverty. Then a liberal (or straw liberal) protests that hard work has nothing to do with it, success is determined by accidents of birth like who your parents are and what your skin color is et cetera, so the poor are blameless in their own predicament.

I’m oversimplifying things, but again the compassionate/sympathetic/progressive side of the debate – and the side endorsed by many of the poor themselves – is supposed to be that success is due to accidents of birth, and the less compassionate side is that success depends on hard work and perseverance and grit and willpower.

The obvious pattern is that attributing outcomes to things like genes, biology, and accidents of birth is kind and sympathetic. Attributing them to who works harder and who’s “really trying” can stigmatize people who end up with bad outcomes and is generally viewed as Not A Nice Thing To Do.

And the weird thing, the thing I’ve never understood, is that intellectual achievement is the one domain that breaks this pattern.

Here it’s would-be hard-headed conservatives arguing that intellectual greatness comes from genetics and the accidents of birth and demanding we “accept” this “unpleasant truth”.

And it’s would-be compassionate progressives who are insisting that no, it depends on who works harder, claiming anybody can be brilliant if they really try, warning us not to “stigmatize” the less intelligent as “genetically inferior”.

I can come up with a few explanations for the sudden switch, but none of them are very principled and none of them, to me, seem to break the fundamental symmetry of the situation. I choose to maintain consistency by preserving the belief that overweight people, depressed people, and poor people aren’t fully to blame for their situation – and neither are unintelligent people. It’s accidents of birth all the way down. Intelligence is mostly genetic and determined at birth – and we’ve already determined in every other sphere that “mostly genetic and determined at birth” means you don’t have to feel bad if you got the short end of the stick.

Consider for a moment Srinivasa Ramanujan, one of the greatest mathematicians of all time. He grew up in poverty in a one-room house in small-town India. He taught himself mathematics by borrowing books from local college students and working through the problems on his own until he reached the end of the solveable ones and had nowhere else to go but inventing ways to solve the unsolveable ones.

There are a lot of poor people in the United States today whose life circumstances prevented their parents from reading books to them as a child, prevented them from getting into the best schools, prevented them from attending college, et cetera. And pretty much all of those people *still* got more educational opportunities than Ramanujan did.

And from there we can go in one of two directions. First, we can say that a lot of intelligence is innate, that Ramanujan was a genius, and that we mortals cannot be expected to replicate his accomplishments.

Or second, we can say those poor people are *just not trying hard enough*.

Take “innate ability” out of the picture, and if you meet a poor person on the street begging for food, saying he never had a chance, your reply must be “Well, if you’d just borrowed a couple of math textbooks from the local library at age 12, you would have been a Fields Medalist by now. I hear that pays pretty well.”

The best reason *not* to say that is that we view Ramanujan as intellectually gifted. But the very phrase tells us where we should classify that belief. Ramanujan’s genius is a “gift” in much the same way your parents giving you a trust fund on your eighteenth birthday is a “gift”, and it should be weighted accordingly in the moral calculus.

II

I shouldn’t pretend I’m worried about this for the sake of the poor. I’m worried for *me*.

My last IQ-ish test was my SATs in high school. I got a perfect score in Verbal, and a good-but-not-great score in Math.

And in high school English, I got A++s in all my classes, Principal’s Gold Medals, 100% on tests, first prize in various state-wide essay contests, etc. In Math, I just barely by the skin of my teeth scraped together a pass in Calculus with a C-.

Every time I won some kind of prize in English my parents would praise me and say I was good and should feel good. My teachers would hold me up as an example and say other kids should try to be more like me. Meanwhile, when I would bring home a report card with a C- in math, my parents would have concerned faces and tell me they were disappointed and I wasn’t living up to my potential and I needed to work harder et cetera.

And *I don’t know which part bothered me more*.

Every time I was held up as an example in English class, I wanted to crawl under a rock and die. I didn’t do it! I didn’t study at all, half the time I did the homework in the car on the way to school, those essays for the statewide competition were thrown together on a lark without a trace of real effort. To praise me for any of it seemed and still seems utterly unjust.

On the other hand, to this day I believe I deserve a fricking *statue* for getting a C- in Calculus I. It should be in the center of the schoolyard, and have a plaque saying something like “Scott Alexander, who by making a herculean effort managed to pass Calculus I, even though they kept throwing random things after the little curly S sign and pretending it made sense.”

And without some notion of innate ability, I don’t know what to do with this experience. I don’t want to have to accept the blame for being a lazy person who just didn’t try hard enough in Math. But I *really* don’t want to have to accept the credit for being a virtuous and studious English student who worked harder than his peers. I *know* there were people who worked harder than I did in English, who poured their heart and soul into that course – and who still got Cs and Ds. To deny innate ability is to devalue their efforts and sacrifice, while simultaneously giving me credit I don’t deserve.

Meanwhile, there were some students who did better than I did in Math with seemingly zero effort. I didn’t begrudge those students. But if they’d started trying to say they had exactly the same level of innate ability as I did, and the only difference was *they* were trying while *I* was slacking off, then I sure as hell would have begrudged them. Especially if I knew they were lazing around on the beach while I was poring over a textbook.

I tend to think of social norms as contracts bargained between different groups. In the case of attitudes towards intelligence, those two groups are smart people and dumb people. Since I was both at once, I got to make the bargain

with myself, which simplified the bargaining process immensely. The deal I came up with was that I wasn't going to beat myself up over the areas I was bad at, but I also didn't get to become too cocky about the areas I was good at. It was all genetic luck of the draw either way. In the meantime, I would try to press as hard as I could to exploit my strengths and cover up my deficiencies. So far I've found this to be a really healthy way of treating myself, and it's the way I try to treat others as well.

III

The theme continues to be "Scott Relives His Childhood Inadequacies". So:

When I was 6 and my brother was 4, our mom decided that as an Overachieving Jewish Mother she was contractually obligated to make both of us learn to play piano. She enrolled me in a Yamaha introductory piano class, and my younger brother in a Yamaha 'cute little kids bang on the keyboard' class.

A little while later, I noticed that my brother was now with me in my Introductory Piano class.

A little while later, I noticed that my brother was now by far the best student in my Introductory Piano Class, even though he had just started and was two or three years younger than anyone else there.

A little while later, Yamaha USA flew him to Japan to show him off before the Yamaha corporate honchos there.

Well, one thing led to another, and my brother won several international piano competitions, got a professorship in music at age 25, and now routinely gets news articles written about him calling him "among the top musicians of his generation".

Meanwhile, I was always a mediocre student at Yamaha. When the time came to try an instrument in elementary school, I went with the violin to see if maybe I'd find it more to my tastes than the piano. I was quickly sorted into the remedial class because I couldn't figure out how to make my instrument stop sounding like a wounded cat. After a year or so of this, I decided to switch to fulfilling my music requirement through a choir, and everyone who'd had to listen to me breathed a sigh of relief.

Every so often I wonder if somewhere deep inside me there is the potential to be "among the top musicians of my generation." I try to recollect whether my brother practiced harder than I did. My memories are hazy, but I don't think he practiced *much* harder until well after his career as a child prodigy had taken off. The cycle seemed to be that every time he practiced, things came fluidly to him and he would produce beautiful music and everyone would be amazed. And this must have felt great, and incentivized him to practice more, and that made him even better, so that the beautiful music came even more fluidly, and the praise became more effusive, until eventually he chose a full-time career in music and became amazing. Meanwhile, when I started practicing it always sounded like wounded cats, and I would get very cautious praise like "Good job, Scott, it sounded like that cat was hurt a little less badly than usual," and it made me frustrated, and want to practice less, which made me even worse, until eventually I quit in disgust.

On the other hand, I know people who want to get good at writing, and make a mighty resolution to write two hundred words a day every day, and then after the first week they find it's too annoying and give up. These people think I'm amazing, and why shouldn't they? I've written a few hundred to a few thousand words pretty much every day for the past ten years.

But as I've said before, this has taken exactly zero willpower. It's more that I can't stop even if I want to. Part of that is probably that when I write, I feel really good about having expressed exactly what it was I meant to say. Lots of people read it, they comment, they praise me, I feel good, I'm encouraged to keep writing, and it's exactly the same virtuous

cycle as my brother got from his piano practice.

And so I think it would be *too easy* to say something like “There’s no innate component at all. Your brother practiced piano really hard but almost never writes. You write all the time, but wimped out of practicing piano. So what do you expect? You both got what you deserved.”

I tried to practice piano as hard as he did. I really tried. But every moment was a struggle. I could keep it up for a while, and then we’d go on vacation, and there’d be no piano easily available, and I would be breathing a sigh of relief at having a ready-made excuse, and he’d be heading off to look for a piano somewhere to practice on. Meanwhile, I am writing this post in short breaks between running around hospital corridors responding to psychiatric emergencies, and there’s probably someone very impressed with that, someone saying “But you had such a great excuse to get out of your writing practice!”

I dunno. But I don’t think of myself as working hard at any of the things I am good at, in the sense of “exerting vast willpower to force myself kicking and screaming to do them”. It’s possible I *do* work hard, and that an outside observer would accuse me of eliding how hard I work, but it’s not a conscious elision and I don’t feel that way from the inside.

Ramanujan worked very hard at math. But I don’t think he thought of it as work. He obtained a scholarship to the local college, but dropped out almost immediately because he couldn’t make himself study any subject other than math. Then he got accepted to another college, and dropped out *again* because they made him study non-mathematical subjects and he failed a physiology class. Then he nearly starved to death because he had no money and no scholarship. To me, this doesn’t sound like a person who just happens to be very hard-working; if he had the ability to study other subjects he would have, for no reason other than that it would have allowed him to stay in college so he could keep studying math. It seems to me that in some sense Ramanujan was *incapable* of putting hard work into non-math subjects.

I really wanted to learn math and failed, but I did graduate with honors from medical school. Ramanujan really wanted to learn physiology and failed, but he did become one of history’s great mathematicians. So which one of us was the hard worker?

People used to ask me for writing advice. And I, in all earnestness, would say “Just transcribe your thoughts onto paper exactly like they sound in your head.” It turns out that doesn’t work for other people. Maybe it doesn’t work for me either, and it just feels like it does.

But you know what? When asked about one of his discoveries, a method of simplifying a very difficult problem to a continued fraction, Ramanujan described his thought process as: “It is simple. The minute I heard the problem, I knew that the answer was a continued fraction. ‘Which continued fraction?’ I asked myself. Then the answer came to my mind”.

And again, maybe that’s just how it feels to him, and the real answer is “study math so hard that you flunk out of college twice, and eventually you develop so much intuition that you can solve problems without thinking about them.”

(or maybe the real answer is “have dreams where obscure Hindu gods appear to you as drops of blood and reveal mathematical formulae”. [Ramanujan was weird](#)

But I *still* feel like there’s something going on here where the solution to me being bad at math and piano isn’t just “sweat blood and push through your brain’s aversion to these subjects until you make it stick”. When I read biographies of Ramanujan and other famous mathematicians, there’s no sense that they ever had to do that with math.

When I talk to my brother, I never get a sense that he had to do that with piano. And if I am good enough at writing to qualify to have an opinion on being good at things, then I don't feel like I ever went through that process myself.

So this too is part of my deal with myself. I'll try to do my best at things, but if there's something I really hate, something where I have to go uphill every step of the way, then it's okay to admit mediocrity. I won't beat myself up for not forcing myself kicking and screaming to practice piano. And in return I won't become too cocky about practicing writing a lot. It's probably [some kind of luck of the draw](#) either way.

IV

I said before that this wasn't just about poor people, it was about me being selfishly worried for my own sake. I think I might have given the mistaken impression that I merely need to justify to myself why I can't get an A in math or play the piano. But it's much worse than that.

The rationalist community tends to get a lot of high-scrupulosity people, people who tend to beat themselves up for not doing more than they are. It's why [I push giving 10% to charity](#), not as some kind of amazing stretch goal that we need to guilt people into doing, but as a crutch, a sort of "don't worry, you're still okay if you only give ten percent". It's why there's so much emphasis on "heroic responsibility" and how you, yes you, have to solve all the world's problems personally. It's why I see red when anyone accuses us of entitlement, since it goes about as well as calling an anorexic person fat.

And we really aren't doing ourselves any favors. For example, Nick Bostrom writes:

Searching for a cure for aging is not just a nice thing that we should perhaps one day get around to. It is an urgent, screaming moral imperative. The sooner we start a focused research program, the sooner we will get results. It matters if we get the cure in 25 years rather than in 24 years: a population greater than that of Canada would die as a result.

If that bothers you, you *definitely* shouldn't read [Astronomical Waste](#).

Yet here I am, not doing anti-aging research. Why not?

Because I tried doing biology research a few times and it was really hard and made me miserable. You know how in every science class, when the teacher says "Okay, pour the white chemical into the grey chemical, and notice how it turns green and begins to bubble," there's always one student who pours the white chemical into the grey chemical, and it just forms a greyish-white mixture and sits there? That was me. I hated it, I didn't have the dexterity or the precision of mind to do it well, and when I finally finished my required experimental science classes I was happy never to think about it again. Even the abstract intellectual part of it – the one where you go through data about genes and ligands and receptors in supercentenarians and shake it until data comes out – requires exactly the kind of math skills that I don't have.

Insofar as this is a matter of innate aptitude – some people are cut out for biology research and I'm not one of them – all is well, and my decision to get a job I'm good at instead is entirely justified.

But insofar as there's no such thing as innate aptitude, just hard work and grit – then by not being gritty enough, I'm a monster who's complicit in the death of a population greater than that of Canada.

Insofar as there's no such thing as innate aptitude, I have *no excuse* for not being Aubrey de Grey. Or if Aubrey de Grey

doesn't impress you much, Norman Borlaug. Or if you don't know who either of those two people are, Elon Musk.

I once heard a friend, upon his first use of modafinil, wonder aloud if the way they felt on that stimulant was the way Elon Musk felt all the time. That tied a lot of things together for me, gave me an intuitive understanding of what it might "feel like from the inside" to be Elon Musk. And it gave me a good tool to discuss biological variation with. Most of us agree that people on stimulants can perform in ways it's difficult for people off stimulants to match. Most of us agree that there's nothing magical about stimulants, just changes to the levels of dopamine, histamine, norepinephrine et cetera in the brain. And most of us agree there's a lot of natural variation in these chemicals anyway. So "me on stimulants is that guy's normal" seems like a good way of cutting through some of the philosophical difficulties around this issue.

...which is all kind of a big tangent. The point I want to make is that for me, what's at stake in talking about natural variations in ability isn't just whether I have to feel like a failure for not getting an A in high school calculus, or not being as good at music as my brother. It's whether I'm a failure for not being Elon Musk. Specifically, it's whether I can say "No, I'm really not cut out to be Elon Musk" and go do something else I'm better at without worrying that I'm killing everyone in Canada.

V

The proverb says: "Everyone has somebody better off than they are and somebody worse off than they are, with two exceptions." When we accept that we're all in the "not Elon Musk" boat together (with one exception) a lot of the status games around innate ability start to seem less important.

Every so often an overly kind commenter here praises my intelligence and says they feel intellectually inadequate compared to me, that they wish they could be at my level. But at my level, I spend my time feeling intellectually inadequate [Book-Review-And-Highlights-Quantum-Computing-Since-Democritus | compared to Scott Aaronson]. Scott Aaronson [describes](#) feeling "in awe" of Terence Tao and frequently struggling to understand him. Terence Tao – well, I don't know if he's religious, but maybe he feels intellectually inadequate compared to God. And God feels intellectually inadequate compared to John von Neumann.

So there's not much point in me feeling inadequate compared to my brother, because even if I was as good at music as my brother, I'd probably just feel inadequate for not being Mozart.

And asking "Well what if you just worked harder?" can elide small distinctions, but not bigger ones. If my only goal is short-term preservation of my self-esteem, I can imagine that if only things had gone a little differently I could have practiced more and ended up as talented as my brother. It's a lot harder for me to imagine the course of events where I do something different and become Mozart. Only one in a billion people reach a Mozart level of achievement; why would it be me?

If I loved music for its own sake and wanted to be a talented musician so I could express the melodies dancing within my heart, then none of this matters. But insofar as I want to be good at music because *I feel bad that other people are better than me at music*, that's a road without an end.

This is also how I feel of when some people on this blog complain they feel dumb for not being as smart as some of the other commenters on this blog.

I happen to have all of your IQ scores in a spreadsheet right here (remember that survey you took?). Not a single person is below the population average. The first percentile for IQ here – the one such that 1% of respondents are lower and 99% of respondents are higher – is – corresponds to the 85th percentile of the general population. So even

if you're in the first percentile here, you're still pretty high up in the broader scheme of things.

At that point we're back on the road without end. I am pretty sure we can raise your IQ as much as you want and you will *still* feel like pond scum. If we raise it twenty points, you'll try reading [Quantum Computing since Democritus](#) and feel like pond scum. If we raise it forty, you'll just go to [Terence Tao's blog](#) and feel like pond scum there. Maybe if you were literally the highest-IQ person in the entire world you would feel good about yourself, but any system where only one person in the world is allowed to feel good about themselves at a time *is a bad system*.

People say we should stop talking about ability differences so that stupid people don't feel bad. I say that there's more than enough room for *everybody* to feel bad, smart and stupid alike, and not talking about it won't help. What will help is fundamentally uncoupling perception of intelligence from perception of self-worth.

I work with psychiatric patients who tend to have cognitive difficulties. Starting out in the Detroit ghetto doesn't do them any favors, and then they get conditions like bipolar disorder and schizophrenia that [actively lower IQ](#) for poorly understood neurological reasons.

The standard psychiatric evaluation includes an assessment of cognitive ability; the one I use is a quick test with three questions. The questions are – “What is 100 minus 7?”, “What do an apple and an orange have in common?”, and “Remember these three words for one minute, then repeat them back to me: house, blue, and tulip”.

There are a lot of people – and I don't mean floridly psychotic people who don't know their own name, I mean ordinary reasonable people just like you and me – who can't answer these questions. And we know why they can't answer these questions, and it is pretty darned biological.

And if our answer to “I feel dumb and worthless because my IQ isn't high enough” is “don't worry, you're not worthless, I'm sure you can be a great scientist if you just try hard enough”, then we are implicitly throwing under the bus all of these people who are *definitely* not going to be great scientists no matter how hard they try. Talking about trying harder can obfuscate the little differences, but once we're talking about the homeless schizophrenic guy from Detroit who can't tell me 100 minus 7 to save his life, you can't just magic the problem away with a wave of your hand and say “I'm sure he can be the next Ramanujan if he keeps a positive attitude!” You either need to condemn him as worthless *or else stop fricking tying worth to innate intellectual ability*.

This is getting pretty close to what I was talking about in my post on [burdens](#). When I get a suicidal patient who thinks they're a burden on society, it's nice to be able to point out ten important things they've done for society recently and prove them wrong. But sometimes it's not that easy, and the only thing you can say is “f#@k that s#!t”. Yes, society has organized itself in a way that excludes and impoverishes a bunch of people who could have been perfectly happy in the state of nature picking berries and hunting aurochs. It's not your fault, and if they're going to give you compensation *you take it*. And we had better make this perfectly clear now, so that when everything becomes automated and run by robots and we're *all* behind the curve, everybody agrees that us continuing to exist is still okay.

Likewise with intellectual ability. When someone feels sad because they can't be a great scientist, it is nice to be able to point out all of their intellectual strengths and tell them “Yes you can, if only you put your mind to it!” But this is often not true. At that point you have to say “f#@k it” and tell them to stop tying their self-worth to being a great scientist. And we had better establish that now, before transhumanists succeed in creating superintelligence and we *all* have to come to terms with our intellectual inferiority.

VI

But I think the situation can also be somewhat rosier than that.

Ozy once told me that the [law of comparative advantage](#) was one of the most inspirational things they had ever read. This was sufficiently strange that I demanded an explanation.

Ozy said that it proves *everyone can contribute*. Even if you are worse than everyone else at everything, you can still participate in global trade and other people will pay you money. It may not be very much money, but it will be some, and it will be a measure of how your actions are making other people better off and they are grateful for your existence.

(in real life this doesn't work for a couple of reasons, but who cares about real life when we have *a theory*?)

After some thought, I was also inspired by this.

I'm never going to be a great mathematician or Elon Musk. But if I pursue my comparative advantage, which right now is medicine, I can still make money. And if I feel like it, I can donate it to mathematics research. Or anti-aging research. Or [the same people Elon Musk donates his money to](#). They will use it to hire smart people with important talents that I lack, and I will be at least partially responsible for those people's successes.

If I had an IQ of 70, I think I would still want to pursue my comparative advantage – even if that was ditch-digging, or whatever, and donate that money to important causes. It might not be very much money, but it would be *some*.

Our modern word "talent" comes from the Greek word *talanton*, a certain amount of precious metal sometimes used as a denomination of money. The etymology passes through a parable of Jesus'. A master calls three servants to him and gives the first five talents, the second two talents, and the third one talent. The first two servants invest the money and double it. The third literally buries it in a hole. The master comes back later and praises the first two servants, but sends the third servant to Hell (metaphor? what metaphor?).

Various people have come up with various interpretations, but the most popular says that God gives all of us different amounts of resources, and He will judge us based on how well we use these resources rather than on how many He gave us. It would be stupid to give your first servant five loads of silver, then your second servant two loads of silver, then immediately start chewing out the second servant for having less silver than the first one. And if both servants invested their silver wisely, it would be silly to chew out the second one for ending up with less profit when he started with less seed capital. The moral seems to be that if you take what God gives you and use it wisely, you're fine.

The modern word "talent" comes from this parable. It implies "a thing God has given you which you can invest and give back".

So if I were a ditch-digger, I think I would dig ditches, donate a portion of the small amount I made, and trust that I had done what I could with the talents I was given. **VII.** The Jews *also* talk about how God judges you for your gifts. Rabbi Zusya once said that when he died, he wasn't worried that God would ask him "Why weren't you Moses?" or "Why weren't you Solomon?" But he did worry that God might ask "Why weren't you Rabbi Zusya?"

And this is part of why it's important for me to believe in innate ability, and especially differences in innate ability. If everything comes down to hard work and positive attitude, then God has every right to ask me "Why weren't you Srinivasa Ramanujan?" or "Why weren't you Elon Musk?"

If everyone is legitimately a different person with a different brain and different talents and abilities, then all God gets to ask me is whether or not I was Scott Alexander.

This seems like a gratifyingly low bar. [*more to come on this subject later*]

I Can Tolerate Anything Except The Outgroup

Posted on September 30, 2014 by Scott Alexander



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 also considered a virtue, but it suffers the same sort of diminished 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 Virtue Points have I earned for my meritorious deeds?"

Bodhidharma answers: "None at all".

The Emperor, somewhat put out, demands to know why.

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 paens to how great the outgroup is, staying up at night fretting that somebody else might not like the outgroup enough.

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 promote and defend their outgroups, the outer the better.

What is going on here?

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 the Nazis and Japanese mostly got along pretty well. Heck, the Nazis were actually moderately positively disposed to 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.

III

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 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.)

IV

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 1/10⁴⁵? 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 sophisticated 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 – 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,)

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.

V

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.

VI

"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.

VII

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*

July's 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.

VIII

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*.

IX

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 so we 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?

X

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 happens 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.

XI

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.

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 embarrass 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*."

Can Things Be Both Popular And Silenced?

Posted on May 23, 2018 by Scott Alexander



The *New York Times* recently reported on various anti-PC thinkers as “[the intellectual dark web](#)”, sparking various annoying discussion.

The first talking point – that the term is silly – is surely true. So is the second point – that it awkwardly combines careful and important thinkers like Eric Weinstein with awful demagogues like Ben Shapiro. So is the third – that people have been complaining about political correctness for decades, so anything that portrays this as a sudden revolt is ahistorical. There are probably more good points buried within the chaff.

But I want to focus on one of the main arguments that’s been emphasized in pretty much every article: can a movement really claim it’s being silenced if it’s actually pretty popular?

“Silenced” is the term a lot of these articles use, and it’s a good one. “Censored” awkwardly suggests government involvement, which nobody is claiming. “Silenced” just suggests that there’s a lot of social pressure on its members to shut up. But shutting up is of course the exact opposite of what the people involved are doing – as the *Times* points out, several IDW members have audiences in the millions, monthly Patreon revenue in the five to six figures, and (with a big enough security detail) regular college speaking engagements.

So, from *New Statesman*, [If The “Intellectual Dark Web” Are Being Silenced, Why Do We Need To Keep Hearing About Them?](#):

The main problem with the whole profile is that it struggles because of a fundamental inherent contradiction in its premise, which is that this group of renegades has been shunned but are also incredibly popular. Either they are persecuted victims standing outside of society or they are not. Joe Rogan “hosts one of the most popular podcasts in the country”, Ben Shapiro’s podcast “gets 15 million downloads a month”. Sam Harris “estimates that his Waking Up podcast gets one million listeners an episode”. Dave Rubin’s YouTube show has “more than 700,000 subscribers”, Jordan Peterson’s latest book is a bestseller on Amazon [...]

On that basis alone, should this piece have been written at all? The marketplace of ideas that these folk are always banging on about is working. They have found their audience, and are not only popular but raking it in via Patreon accounts and book deals and tours to sold-out venues. Why are they not content with that? They are not content with that because they want everybody to listen, and they do not want to be challenged.

In the absence of that, they have made currency of the claim of being silenced, which is why we are in this ludicrous position where several people with columns in mainstream newspapers and publishing deals are going around with a loudhailer, bawling that we are not listening to them.

[Reason](#)’s article is better and makes a lot of good points, but it still emphasizes this same question, particularly in their subtitle: “The leading figures of the ‘Intellectual Dark Web’ are incredibly popular. So why do they still feel so aggrieved?”. From the piece:

They can be found gracing high-profile cable-news shows, magazine opinion pages, and college speaking tours. They've racked up hundreds of thousands of followers. And yet the ragtag band of academics, journalists, and political pundits that make up the "Intellectual Dark Web" (IDW)—think of it as an Island of Misfit Ideologues—declare themselves, Trump-like, to be underdogs and outsiders. [...]

[I'm not convinced] they're actually so taboo these days. As Weiss points out, this is a crowd that has built followings on new-media platforms like YouTube and Twitter rather than relying solely on legacy media, academic publishing, and other traditional routes to getting opinions heard. (There isn't much that's new about this except the media involved. Conservatives have long been building large audiences using outside-the-elite-media platforms such as talk radio, speaking tours, and blogs.) In doing so, they've amassed tens and sometimes hundreds of thousands of followers. What they are saying might not be embraced, or even endured, by legacy media institutions or certain social media precincts, but it's certainly not out of tune with or heretical to many Americans.

The bottom line is there's no denying most of these people are very popular. Yet one of the few unifying threads among them is a feeling or posture of being marginalized, too taboo for liberal millennial snowflakes and the folks who cater to them.

The basic argument – that you can't be both silenced and popular at the same time – sounds plausible. But I want to make a couple points that examine it in more detail.

1. There are lots of other cases where we would agree there's some form of silencing going on, even as a group has many supporters and rich, famous spokespeople

I know a lot of closeted transgender people. They're afraid to come out as trans, they talk about trans people being stigmatized and silenced, and they clearly have a point. Does anyone disagree that it can be dangerous to be a trans person even in the First World, let alone anywhere else?

On the other hand, Caitlyn Jenner is on the cover of every magazine, won Woman Of The Year, got her own documentary and reality TV show, and earns [up to \\$100,000](#) per public appearance, with a [total net worth](#) rumored to be around \$100 million. She is probably one of the most famous and popular people in the world.

Only a moron would make an argument like "Caitlyn Jenner is doing very well, therefore there's not really a stigma around transgender". For one thing, your success is a function of how many people like you, not your net (likers – haters) total. For another, Hollywood is its own world and probably doesn't correlate with any particular person's social sphere. And for another, Jenner is popular partly because of how surprising and controversial her transition was – her story is at least partly a function of "look how brave this person is to defy social stigma this way".

Transgender people complain of social shaming, silencing, and stigma. Some transgender people can become very famous celebrities who everyone agrees are rich and popular. And nobody finds this at all surprising or thinks that these two claims contradict each other.

(No, Twitter, I'm not making the claim "Sam Harris is exactly as marginalized as transgender people". I'm saying that even groups who we all agree are more marginalized than the IDW can have very successful and famous spokespeople.)

Or what about the early US labor movement? They were faced with everything from Pinkerton goon squads, to industry blacklists, to constantly getting arrested on trumped-up charges; nobody seriously denies that government and private industry put a lot of effort into silencing them.

Yet they were very popular with their core demographic, and their most charismatic spokespeople remained famous and widely-liked. Emma Goldman would go around the country lecturing to packed halls, collecting far more energy and interest than Sam Harris gets nowadays when he does the same. If the papers of the time had said “Emma Goldman sure is popular for someone who says her movement is being silenced”, well, screw you and your dumb gotchas, that’s just a 100% accurate description of the state of affairs.

2. In fact, taboo opinions seem to promote a culture of celebrity

From [Current Affairs](#):

There are dozens of well-known critics of social justice activists: Harris, Shapiro, Peterson, Brooks, Stephens, Hoff Sommers, Weinstein, Weinstein, Murray, Murray, Rogan, Chait, Haidt, Pinker, Rubin, Sullivan, Weiss, Williamson, Yiannopoulos, Dreger, Hirsi Ali. Who are their equivalents among the Social Justice Types? Who has their reach or prominence?

A few people have tried to answer the question – and certainly a few names like Ta-Nehisi Coates belong in any such list. But I think the overall point is basically correct. If so, what does that mean?

Consider this: how many neo-Nazi/white supremacist activists are famous enough that the average news junkie would know their names? Maybe two: David Duke and Richard Spencer. Okay. How many low-tax activists are equally famous? I think just one: Grover Norquist. There are some important people who happen to support low taxes among many other causes (eg Paul Ryan) but they don’t count – if they did, our list of famous “social justice types” would have to include Hillary Clinton and a hundred others.

Presumably we shouldn’t conclude that neo-Nazism is twice as common/popular/acceptable as tax cuts. But that means you can’t always measure how popular an ideology is by counting its famous advocates.

I’d go further and say that more taboo ideas are more likely to generate famous spokespeople. If you can’t think of any modern feminists with star power, you can always go back to the 1970s and find people like Gloria Steinem and Andrea Dworkin – who made waves by being at least as outrageous then as the IDW is now. If Ta-Nehisi Coates isn’t famous enough for you, Martin Luther King and Malcolm X certainly will be. Malcolm X didn’t get more famous than Coates by being *more well-liked*, he got famous by being as controversial and threatening and feared as Coates is accepted. So the implication of the *Current Affairs* article – we mostly hear about well-liked people, and really controversial people never get famous – seems questionable at best and backwards at worse.

But why would more taboo causes generate more celebrity? Here are some ways I think this could work:

1. Controversy sells in general. Caitlyn Jenner is more famous than Bruce Jenner not because transgender is less stigmatized than running, but because it’s more likely to provoke debate.
2. All else being equal, if an ideology is taboo, it should have fewer loud open activists per covert believer than an orthodox ideology. But that means the field is less crowded. If feminism has 1 loud activist per 10 believers, and the IDW has 1 loud activist per 1000 believers, then the feminist activist will generally be speaking to a college club, and the IDW activist to a crowded lecture hall. This will catapult the IDW activists to greater celebrity.
3. Activists for taboo views need a skill that activists for orthodox views don’t – that of surfing controversy. The insult “edgelord” is basically correct – they thrive by being on the edge of what is acceptable. If you go completely beyond the bounds of what is acceptable, you fall from grace – either into literal ruin, or just having your fan base shift entirely to being weird alt-right people whom you hate and don’t want to be associated with. Only people who can continually surf that boundary – edgy enough to be interesting, restrained enough to get

the *New York Times* to write basically positive editorials about you – are really able to make it. Most people correctly assume they would screw up and end up totally taboo rather than delightfully edgy. Once again, this makes the field less crowded, giving everyone who comes in more star power per person.

4. Orthodox ideologies tend to be well-represented within institutions, meaning that the ideologies' leaders are more likely to be institutionally prestigious people. Taboo views are unrepresented within institutions, meaning their spokespeople kind of just arise naturally by being really good at getting attention and acclaim. The natural “leaders of feminism” might be Women’s Studies professors, Planned Parenthood directors, and whoever the most feminist person at the *New York Times* is. These people might be very good at what they do, they might even be very effective at promoting feminism, but they’re probably less good at getting attention than people who have been specifically selected for that trait. And with the institutional leaders sucking up all the status, it might be harder for some woman who’s just a very good writer and really in-touch with the zeitgeist to say “Yes, I am the leader of feminism, everyone please care about me now”.
5. Generic famous people will support orthodox causes, but not taboo causes. The absence of people famous for feminism is counterbalanced by a glut of famous people who happen to be feminists. Here is a [list of actors who say they are proud to call themselves feminist](#), also just known as “a list of actors”. Famous people who are against feminism are more likely to keep quiet about it, creating a void for specific anti-feminist celebrities can fill.
6. Celebrity helps launder taboo ideology. If you believe Muslim immigration is threatening, you might not be willing to say that aloud – especially if you’re an ordinary person who often trips on their tongue, and the precise words you use are the difference between “mainstream conservative belief” and “evil bigot who must be fired immediately”. Saying “I am really into Sam Harris” both leaves a lot of ambiguity, and lets you outsource the not-saying-the-wrong-word-and-getting-fired work to a professional who’s good at it. In contrast, if your belief is orthodox and you expect it to win you social approval, you want to be as direct as possible.

I don’t know if these six points really explain the phenomenon. But I think there’s definitely a phenomenon to be explained, and I think “crowded field” is a big part of it. In my own experience, my blog posts promoting orthodox opinions are generally ignored; my blog posts promoting controversial opinions go viral and win me lots of praise. I assume this is because my orthodox blog posts are trying to outcompete the people at Vox (highly-polished, Ivy-League-educated mutants grown in vats by a DARPA project to engineer the perfect thinkpiece writer), and my controversial blog posts are trying to outcompete three rando blogs that consistently confuse “there” and “their”. Winning one competition is much easier than winning the other – and the prize for winning either is “the attention of about 50% of the population”.

3. Fame lets people avoid social repercussions, but that doesn’t mean those repercussions don’t exist for ordinary people

Caitlyn Jenner can be as visibly and fabulously transgender as she wants, because being transgender is a big part of her job. She’s organized a lot of her life around being a transgender person. Any friends she was going to lose for being transgender have already been written off as losses. Anybody who wants to harm her for being transgender is going to get stopped by her bodyguards or kept out of her giant gated mansion. When she argues that transgender people face a lot of stigma, fear, and discrimination, she mostly isn’t talking about herself. She’s talking about every transgender person who isn’t Caitlyn Jenner.

Likewise, Sam Harris is pretty invincible. As a professional edgelord, he is not going to lose his job for being edgy. Whatever friends he’s going to lose for being Sam Harris, he’s already written off as losses. I assume he has some kind of security or at least chooses not to live in Berkeley. So when he’s talking about his ideas being taboo, he means *taboo for everybody who isn’t Sam Harris*.

I worry that this conversation is being conducted mostly by media personalities who write controversial takes for a living. They work for ideologically-aligned publications, and everyone knows that a few crazies hating and harassing you is a common part of the job. If you didn't propose the death penalty for abortion and then get a job at *The Atlantic*, you'll probably be fine.

Out in the rest of the world, if a rando on social media calls your company and tells them you're a Nazi because [out of context tweet], the complaint is going straight to a humorless 60-year-old HR drone whose job is minimizing the risk of PR blowups, and who has never heard of Twitter except as a vague legend of a place where everything is terrible all the time. So if you write for a webzine, consider that you may have no idea how silenced or living-in-fear anyone else is or isn't, and that you may be the wrong person to speculate about it.

Out in the rest of the world, if someone sends you a death threat, you might not be such an experienced consumer of Internet vitriol that you know it usually doesn't pan out. You might not be so thick-skinned that "Go to hell, you fucking Nazi scum" no longer has any effect on you. You might not live in a bubble of intellectualism where people appreciate subtle positions. You might have friends and family who are very nice people but somewhat literal-minded, who have heard that only rapists oppose feminism so many times that they have no ability to create a mental category for someone who opposes feminism but isn't pro-rape. And you might not really relish the idea of having to have a conversation with your sweet elderly great-aunt about how no, you *really* don't think raping people is good. Seriously, imagine having to explain *any* of what you write on the Internet to your sweet elderly great-aunt, and now imagine it's something that society has spent years telling her is equivalent to rape apologism.

(my father recently implied I had brought dishonor upon our family by getting quoted approvingly in *National Review*. I am 90% sure he was joking, but only 90%.)

Or maybe I'm wrong about this. Part of how silencing works is that nobody really knows how strong it is or isn't. I had a patient who agonized for years over whether to come out to his family, only to have his parents say "Yes, obviously" when he finally got up the nerve. The point, is *Sam Harris no longer has to worry about any of these things*. So if your line of reasoning is "well, Sam Harris seems to do pretty well for himself, so I guess you can't get in trouble for being controversial", I don't know what to tell you.

4. If you spend decades inventing a powerful decentralized network to allow unpopular voices to be heard, sometimes you end up with unpopular voices being heard

Sam Harris' business model is a podcast with a Patreon, advertised by Internet word-of-mouth. This is pretty typical for the "intellectual dark web" figures.

The Internet promised to take power away from media gatekeepers and make censorship near-impossible. In discussing the many ways in which this promise has admittedly failed, we tend to overlook the degree to which it's succeeded. One of the most common historical tropes is "local government and/or lynch mob destroys marginalized group's printing press to prevent them from spreading their ideas". The Internet has since made people basically uncensorable, not for lack of trying. More recently, crowdfunding has added the final part to this machine – semi-decentralized cash flow.

So, after hundreds of engineers and activists and entrepreneurs work for decades to create a new near-impossible-to-censor system, and some people who would never have gotten heard on any other channel are able to use it to get heard – well, it's pretty weird to turn around and say "Aha, you got popular, that proves nobody is trying to silence you!"

I think this also explains why, even though people have been talking about these issues forever, it's only becoming a "big deal" now. Before, people would either watch their mouths to avoid getting kicked out by major gatekeeper

institutions – or they would go to explicitly right-coded spaces like talk radio where the gatekeepers already agreed with them.

What's new is that there's a third route in between "tame enough to be on CNN" and "conservative enough to be a guest on Rush Limbaugh". The new brand of IDW thinkers are interesting precisely because – excluding Ben Shapiro (always a good life choice) – they're *not* traditional conservatives. The thing that's new and exciting enough to get *New York Times* articles written about it is that the anti-PC movement has spread to friendly coastal liberals. From the Democrats' perspective, the [IDW aren't infidels, they're heretics](#).

5. When the IDW claims they are threatened, harassed, and blacklisted, people should at least consider that they are referring to the actual well-known incidents of threats, harassment, and blacklisting against them rather than imagining this is code for "they demand to be universally liked"

Here are some of the stories in Weiss' original IDW editorial:

A year ago, Bret Weinstein and Heather Heying were respected tenured professors at Evergreen State College, where their Occupy Wall Street-sympathetic politics were well in tune with the school's progressive ethos. Today they have left their jobs, lost many of their friends and endangered their reputations. All this because they opposed a "Day of Absence," in which white students were asked to leave campus for the day. For questioning a day of racial segregation cloaked in progressivism, the pair was smeared as racist. Following threats, they left town for a time with their children and ultimately resigned their jobs.

And:

Mr. Peterson has endured no small amount of online hatred and some real-life physical threats: In March, during a lecture at Queen's University in Ontario, a woman showed up with a garrote.

And:

Dr. Soh said that she started "waking up" in the last two years of her doctorate program. "It was clear that the environment was inhospitable to conducting research," she said. "If you produce findings that the public doesn't like, you can lose your job."

When she wrote an op-ed in 2015 titled "Why Transgender Kids Should Wait to Transition," citing research that found that a majority of gender dysphoric children outgrow their dysphoria, she said her colleagues warned her, "Even if you stay in academia and express this view, tenure won't protect you."

And:

The University of California, Berkeley, had to spend \$600,000 on security for Mr. Shapiro's speech there.

So. Threats against a professor and his family forcing him to leave town. Another professor told that she would lose her job if she communicated research to the public. A guy needing \$600,000 worth of security just to be able to give a speech without getting mobbed. Someone showing up to a lecture with a garrote. And Reason Magazine reads all this and thinks "I know what's going on! These people's only possible complaint is that they feel entitled to have everyone

agree with them!"

Maybe I'm being mean here? But how else do I interpret paragraphs like this one?

The supposed ostracism they suffer because of their views ultimately comes down to a complaint not about censorship or exclusion but being attacked, challenged, or denied very particular opportunities. They want to say the things they are saying and have the marketplace of ideas and attention not only reward them with followers and freelance writing gigs but universal acceptance from those that matter in the academy and chattering classes.

I am nowhere near these people either in fame or controversialness, but I have gotten enough threats and harassment both to be pretty sure that these people are telling the truth, and to expect that the stuff that fits in one article is probably just the tip of the iceberg.

(Do other groups face similar pressures? Absolutely. Would people who wrote similar articles using those groups' complaints to make fun of them also be antisocial? Absolutely.)

On a related note, what does the article mean by contrasting "excluded" vs. "denied very particular opportunities"? I understand the meaning of the words, but I am not sure the people writing about them have a principled distinction in mind. When Debra Soh faced pressure to quit academia, was she being "excluded" or "denied a very particular opportunity"? Would the 1950s version of *Reason* describe communist sympathizers as being "excluded", or as "denied very particular opportunities" in the film industry? If, [as the surveys suggest](#), 20% of philosophers would refuse to hire transgender professors to their department, are transgender people facing "exclusion", or just being "denied very particular opportunities"?

[My position – if you decide not to hire someone based on any characteristic not related to job performance (very broadly defined, including things like company fit and fun to work with), you're trying to exclude people. If you make up a really strained dumb argument for why some characteristic relates to job performance when it obviously doesn't ("communist actors could try to hold a revolution on the set, thus making our other employees feel unsafe"), then you're trying to exclude people *and* lying about it. You can say, as many throughout history have "I'm proud to be part of the effort to fight the Communist menace by denying them positions of influence", and then you get points for honesty and (if the Communists were really as menacing as you thought) maybe utilitarianism points as well. But don't say "What? Me exclude Communists? We're just denying them very particular opportunities! Sure are a whiny bunch, those commies!" See also [Is It Possible To Have Coherent Principles Around Free Speech Norms?](#)]

6. The IDW probably still censor themselves

Another common point in this discussion has been that the IDW copies the worst parts of social justice – intense focus on the latest outrage, shoddy science, its own set of insults ("snowflake! triggered millennial!"), us-vs-them dichotomy, et cetera. And Despite Their Supposed Interest In Rational Discussion Actually They Are Very Bad At Supporting Their Points Rationally.

Here's a site that hasn't been in any "intellectual dark web" editorials and never will be: Human Varieties. You can Google it if you want, but I won't direct-link them for the same way I wouldn't build a giant superhighway to some remote forest village enjoying its peaceful isolation. Here's an excerpt from a typical Human Varieties article:

I did look through the PING survey (age 3-21, N ~ 1,500) – which might not be very informative owing to the age structure. Going by this, Greg [Cochran] seems to be more or less correct about some of the endo[phenotypic] differences and probably about their origins. As an example, Figure 1 & 2 show the

[black/white] diff[erences] for intracranial and total brain volume by age. ([African-Americans] are picked out for illustration since they are the largest non-White ethnic group, showing the biggest deviation from Whites.) And Figure 3 shows the relation between brain volume and ancestry in the self-identified [African-American] group; the results were basically the same for intracranial volume, etc. — and so not shown.

Read Human Varieties for a while, and you notice a few things:

1. They're much more taboo and openly racist (in the Charles Murray [sense](#)) than almost anyone in the "intellectual dark web"
2. They are much less annoying and less likely to shout "TRIGGERED! SNOWFLAKE!" than almost anyone in the "intellectual dark web"
3. Nobody pays any attention to them at all

I think all three of these are correlated.

If you want to be Human Varieties, you can talk about the evidence for and against various taboo subjects. But nobody wants to be them, for two reasons.

First, somebody is going to have to present the evidence for the taboo subject, not just in an edgy "what if... perhaps this should not be suppressed?? or did i blow your mind??" way, but in a "here's exactly what I believe and why I believe it" way. This isn't just Sam Harris level edgy, this is way off the edge into the void below.

Second, if you do even a moderately good job, it's probably going to sound exactly like the quote above, stuff like "this survey of intracranial volume endophenotypes might not be very informative, owing to the age structure" – and everyone will fall asleep by minute two. People will do lots of things to own the libs, but reading an analysis of the age structure of endophenotype data probably isn't one of them.

"TRIGGERED! SNOWFLAKE!" solves both these problems. You avoid the object-level debate about whether taboo subjects are true, and it's automatically interesting to a wide range of people. "That other monkey has status that should be *my* status!" – nobody ever went broke peddling that.

I think this model knocks down a few reasonable-sounding but on-reflection-wrong critiques of the way these issues are discussed:

"The IDW demands rational debate, but they never engage in it". Somewhat true. If they engaged in it, they would move beyond the bounds of acceptable edginess. "We wish we were allowed to talk about X without massive risk to our reputations and safety" and "We are definitely not going to talk about X right now" are hardly contradictory; they follow naturally from each other. And I think this is more subtle than people expect – somebody may feel they can get away with making some arguments but not others, giving them the appearance of a skeletal but flimsy ideology that falls down on close examination. Or people might be willing to talk about these issues in some low-exposure spaces but not other higher-exposure spaces, giving them the appearance of backing down once challenged.

"The IDW focuses too much on triggered snowflakes." Somewhat true – even independent of this being popular and lucrative. This is the least taboo thing you can do while still getting a reputation for being edgy. And winning the free speech wars makes it easier to talk about other stuff.

"The IDW says they're being silenced, but actually they're popular". Somewhat true, even independent of all the arguments above. The things they complain about not being able to say, aren't the things they're saying.

7. Nobody in this discussion seems to really understand how silencing works

If you say “We know a movement isn’t being silenced because it’s got lots of supporters, is widely discussed, and has popular leaders” – then you’re mixing up the numerator and the denominator.

Gandhi’s Indian independence movement had lots of supporters, was widely discussed, and had popular leaders. So did a half dozen Irish revolts against British rule. And the early US labor movement. And Eastern Bloc countries’ resistance to Soviet domination. And Aung San Suu Kyi. And every medieval peasants’ revolt ever. And... well, every other movement that’s been suppressed. Really, what sort of moron wastes their time suppressing a leaderless movement that nobody believes in or cares about?

Popular support and frequent discussion go in the numerator when you’re calculating silencing. Silencing is when *even though* a movement has lots of supporters, none of them will admit to it publicly under their real name. *Even though* a movement is widely discussed, its ideas never penetrate to anywhere they might actually have power. *Even though* it has charismatic leaders, they have to resort to low-prestige decentralized people-power to get their message across, while their opponents preach against them from the airwaves and pulpits and universities.

Scott Aaronson [writes about](#) the game theoretic idea of “common knowledge” as it applies to society:

If you read accounts of Nazi Germany, or the USSR, or North Korea or other despotic regimes today, you can easily be overwhelmed by this sense of, “so why didn’t all the sane people just rise up and overthrow the totalitarian monsters? Surely there were more sane people than crazy, evil ones. And probably the sane people even knew, from experience, that many of their neighbors were sane—so why this cowardice?” Once again, it could be argued that common knowledge is the key. Even if everyone knows the emperor is naked; indeed, even if everyone knows everyone knows he’s naked, still, if it’s not common knowledge, then anyone who says the emperor’s naked is knowingly assuming a massive personal risk. That’s why, in the story, it took a child to shift the equilibrium. Likewise, even if you know that 90% of the populace will join your democratic revolt provided they themselves know 90% will join it, if you can’t make your revolt’s popularity common knowledge, everyone will be stuck second-guessing each other, worried that if they revolt they’ll be an easily-crushed minority. And because of that very worry, they’ll be correct!

(My favorite Soviet joke involves a man standing in the Moscow train station, handing out leaflets to everyone who passes by. Eventually, of course, the KGB arrests him—but they discover to their surprise that the leaflets are just blank pieces of paper. “What’s the meaning of this?” they demand. “What is there to write?” replies the man. “It’s so obvious!” Note that this is precisely a situation where the man is trying to make common knowledge something he assumes his “readers” already know.)

The kicker is that, to prevent something from becoming common knowledge, all you need to do is censor the common-knowledge-producing mechanisms: the press, the Internet, public meetings. This nicely explains why despots throughout history have been so obsessed with controlling the press, and also explains how it’s possible for 10% of a population to murder and enslave the other 90% (as has happened again and again in our species’ sorry history), even though the 90% could easily overwhelm the 10% by acting in concert. Finally, it explains why believers in the Enlightenment project tend to be such fanatical absolutists about free speech.

One can [take this further](#):

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 they don't, everyone else will kill them, and so on.

Suppose in the dictatorless dystopia, one guy becomes immortal for some reason. He goes around saying "Maybe we shouldn't all shock ourselves all the time." Everyone tries to kill him and fails. But if anybody else starts agreeing with him – "Yeah, that guy has a point!" – then everybody kills that other guy.

The don't-shock-ists have 100% popular support. And they have charismatic leaders who get their points out well. But they're still being silenced, and they're still the losing side. Social censorship isn't about your support or your leaders. It's about creating systems of common knowledge that favor your side and handicap your opponents. Censorship = support / common knowledge of support.

Bret Weinstein said of his conflicts with Evergreen State: "I've received... quite a bit of support privately from within the college. Publicly, only one other professor has come forward to say he supports my position." Freddie deBoer [writes about](#) how his own conflicts with callout culture have ended the same way: an outpouring of private emails voicing agreement, plus an outpouring of public comments voicing hostility, *sometimes from the same people privately admitting they agree with him* This provides context for interpreting the *Reason* article's last paragraph:

They want not so much any particular policy platform, political idea, or candidate to catch on as for more people to acknowledge that they are right. And that will always be a proposition that winds up making one feel aggrieved, because it's an impossible one. To the extent that they are spouting marginalized or unpopular ideas, the only way to spread these into the mainstream is to put in the hard work of winning people over.

This is the equivalent of going to communist Czechoslovakia and thinking "Look at all those [greengrocers](#) with communist slogans in their shop windows! Clearly communists have won the war of ideas, and anti-communists are just too aggrieved to do the hard work of convincing people". The other interpretation is that lots of people are already convinced and afraid to say so, and that convincing more people is less productive than building common knowledge of everyone's convictions (maybe you should hand out blank leaflets). I'm not saying convincing people isn't good and necessary, just that assessing how convinced people are is harder than it looks.

Here is a story I heard from a friend, which I will alter slightly to protect the innocent. A prestigious psychology professor signed an open letter in which psychologists condemned belief in innate sex differences. My friend knew that this professor believed such differences existed, and asked him why he signed the letter. He said that he expected everyone else in his department would sign it, so it would look really bad if he didn't. My friend asked why he expected everyone else in his department to sign it, and he said "Probably for the same reason I did".

This is the denominator of silencing in a nutshell. I think it's a heck of a lot more relevant to this discussion than how many Patreon followers Sam Harris has, and I'm happy there are people speaking out against it and trying to make common knowledge a little bit more common.

Social Censorship: The First Offender Model

Posted on April 2, 2019 by Scott Alexander



RJ Zigerell (h/t Marginal Revolution) studies [public support for eugenics](#). He finds that about 40% of Americans support some form of eugenics. The policies discussed were very vague, like “encouraging poor criminals to have fewer children” or “encouraging intelligent people to have more children”; they did not specify what form the encouragement would take. Of note, much lack of support for eugenics was a belief that it would not work; people who believed the qualities involved were heritable were much more likely to support programs to select for them. For example, of people who thought criminality was completely genetic, a full 65% supported encouraging criminals to have fewer children.

I was surprised to hear this, because I thought of moral opposition to eugenics was basically universal. If a prominent politician tentatively supported eugenics, it would provoke a media firestorm and they would get shouted down. This would be true even if they supported the sort of generally mild, noncoercive policies the paper seems to be talking about. How do we square that with a 40% support rate?

I think back to a metaphor for norm enforcement I used in an [argument](#) against Bryan Caplan:

Imagine a town with ten police officers, who can each solve one crime per day. Left to their own devices, the town’s criminals would commit thirty muggings and thirty burglaries per day (for the purposes of this hypothetical, both crimes are equally bad). They also require different skills; burglars can’t become muggers or vice versa without a lot of retraining. Criminals will commit their crime only if the odds are against them getting caught – but since there are 60 crimes a day and the police can only solve ten, the odds are in their favor.

Now imagine that the police get extra resources for a month, and they use them to crack down on mugging. For a month, every mugging in town gets solved instantly. Muggers realize this is going to happen and give up.

At the end of the month, the police lose their extra resources. But the police chief publicly commits that from now on, he’s going to prioritize solving muggings over solving burglaries, even if the burglaries are equally bad or worse. He’ll put an absurd amount of effort into solving even the smallest mugging; this is the hill he’s going to die on.

Suppose you’re a mugger, deciding whether or not to commit the first new mugging in town. If you’re the first guy to violate the no-mugging taboo, every police officer in town is going to be on your case; you’re nearly certain to get caught. You give up and do honest work. Every other mugger in town faces the same choice and makes the same decision. In theory a well-coordinated group of muggers could all start mugging on the same day and break the system, but muggers aren’t really that well-coordinated.

The police chief’s public commitment solves mugging without devoting a single officer’s time to the problem, allowing all officers to concentrate on burglaries. A worst-crime-first enforcement regime has 60 crimes per day and solves 10; a mugging-first regime has 30 crimes per day and solves 10.

But this only works if the police chief keeps his commitment. If someone tests the limits and commits a

mugging, the police need to crack down with what looks like a disproportionate amount of effort – the more disproportionate, the better. Fail, and muggers realize the commitment was fake, and then you're back to having 60 crimes a day.

I think eugenics opponents are doing the same thing as the police here: they're trying to ensure certainty of punishment for the first offender. They've established a norm of massive retaliation against the first person to openly speak out in favor of eugenics, so nobody wants to be the first person. If every one of the 40% of people who support eugenics speak out at once, probably they'll all be fine. But they don't, so they aren't.

Why aren't we in the opposite world, where the people who support eugenics are able to threaten the people who oppose it and prevent them from speaking out? I think just because the opponents coordinated first. In theory one day we could switch to the opposite equilibrium.

I think something like this happened with gay rights. In c. 1969, people were reluctant to speak out in favor of gay rights; in 2019, people are reluctant to speak out against them. Some of that is genuinely changed minds; I don't at all want to trivialize that aspect. But some of it seems to have just been that in 1969, it was common knowledge that the anti-gay side was well-coordinated and could do the massive-retaliation thing, and now it's common knowledge that the pro-gay side is well-coordinated and can do the massive retaliation thing. The switch involved a big battle and lots of people massively retaliating against each other, but it worked.

Maybe everyone else already realized something like this. But it changes the way I think about censorship. I'm still against it. But I used to have an extra argument against it, which was something like "If eugenics is taboo, that means there must be near-universal opposition to eugenics, which means there's no point in keeping it taboo, because even if it wasn't taboo eugenicists wouldn't have any power." I no longer think that argument holds water. "Taboo" might mean nothing more than "one of two equally-sized sides has a tenuous coordination advantage".

(in retrospect I was pretty dumb for not figuring this out, since it's pretty the same argument I make in [Can Things Be Both Popular And Silenced?](#) The answer is obviously yes – if Zigerell's paper is right, eugenics is both popular and silenced – but the police metaphor explains how.)

The strongest argument against censorship is still that beliefs should be allowed to compete in a marketplace of ideas. But if I were pro-censorship, I might retort that one reason to try to maintain my own side's tenuous coordination advantage is that if I relax even for a second, the other side might be able to claw together its own coordination advantage and censor *me*. This isn't possible in the "one side must be overwhelmingly more powerful" model of censorship, but it's something that the "tenuous coordination advantage" model has to worry about. The solution would be some sort of stable structural opposition to censorship in general – but the gay rights example shows that real-world censors can't always expect that to work out for them.

In order to make moderation easier, please restrict yourself to comments about censorship and coordination, not about eugenics or gay rights.

Vote On Values, Outsource Beliefs

Posted on May 8, 2014 by Scott Alexander



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.

Weak Men Are Superweapons

Posted on May 12, 2014 by Scott Alexander



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](http://notallmenarelikethat.tumblr.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.

How Did New Atheism Fail So Miserably?

Posted on October 24, 2017 by Scott Alexander



The Baffler publishes [a long article](#) against “idiot” New Atheists. It’s interesting only in the context of so many [similar articles](#), and an inability to imagine the opposite opinion showing up in an equally fashionable publication. New Atheism has lost its battle for the cultural high ground. r/atheism will shamble on as some sort of undead abomination, chanting “BRAAAAAAAIIINSSSS... are what fundies don’t have” as the living run away shrieking. But everyone else has long since passed them by.

The New Atheists accomplished the seemingly impossible task of alienating a society that agreed with them about everything. The Baffler-journalists of the world don’t believe in God. They don’t disagree that religion contributes to homophobia, transphobia, and the election of some awful politicians – and these issues have only grown more visible in the decade or so since New Atheism’s apogee. And yet in the bubble where nobody believes in God and everyone worries full-time about sexual minorities and Trump, you get less grief for being a Catholic than a Dawkins fan. When Trump wins an election on the back of evangelicals, and the alt-right is shouting “DEUS VULT” and demanding “throne and altar conservatism”, the *real* scandal is rumors that some New Atheist might be reading /pol/. How did the New Atheists become so loathed so quickly?

The second article presents a theory:

It has something to do with a litany of grievances against the believoisie so rote that it might well (or ironically) be styled a catechism. These New Atheists and their many fellow travelers all share an unpleasant obsessive tic: they mouth some obvious banality—there is no God, the holy books were all written by human beings—and then act as if it is some kind of profound insight. This repetition-compulsion seems to be baked right into their dogma.

It compares New Atheists to Kierkegaard’s lunatic:

Soren Kierkegaard, the great enemy of all pedants, offers a story that might shed considerable light. In his Concluding Unscientific Postscript, he describes a psychiatric patient who escapes from the asylum, climbing out a window and running through the gardens to rejoin the world at large. But the madman worries: out in the world, if anyone discovers that he is insane, he will instantly be sent back. So he has to watch what he says, and make sure none of it betrays his inner imbalance—in short, as the not-altogether unmad Danish genius put it, to “convince everyone by the objective truth of what he says that all is in order as far as his sanity is concerned.” Finding a skittle-bowl on the ground and popping it in his pocket, he has an ingenious idea: who could possibly deny that the world is round? So he goes into town and starts endlessly repeating that fact, proffering it over and over again as he wanders about with his small furious paces, the skittle-bowl in his coat clanking, in strict conformity with Newton’s laws, against what Kierkegaard euphemistically refers to as his “a-.” Of course, the poor insistent soul is then sent right back to the asylum [...]

Kierkegaard’s villagers saw someone maniacally repeating that the world is round and correctly sent him back to the asylum. We watched [Neil de Grasse] Tyson doing exactly the same thing, and instead of hiding him away from society where nobody would have to hear such pointless nonsense, thousands cheer him on

for fighting for truth and objectivity against the forces of backwardness. We do the same when Richard Dawkins valiantly fights for the theory of evolution against the last hopeless stragglers of the creationist movement, with their dinky fiberglass dinosaurs munching leaves in a museum-piece Garden of Eden. We do it when Sam Harris prises deep into the human brain and announces that there's no little vacuole there containing a soul.

So the problem with New Atheism was that its whole shtick was repeating obviously true things that everyone already knew? But about 80% of Americans identify as religious, 63% claim to be "absolutely certain" that there is a God, and 46% think the world was literally created in seven days. This is a surprising number of people disagreeing with a thing that everybody already knows.

I could be misreading the article. The article could be wrong. But I don't think so. This is *my* intuitive feeling of what was wrong with New Atheism as well. It wasn't that they were wrong. Just that they were right in a loud, boring, and pointless way.

A charitable reading: New Atheists weren't reaching their intellectual opponents. They were coming into educated urban liberal spaces, saying things that educated urban liberals already believed, and demanding social credit for it. Even though 46% of America is creationist, zero percent of my hundred-or-so friends are. If New Atheists were preaching evolution in social circles like mine, they were wasting their time.

This seems like an accurate criticism of New Atheism, one that earns them all the condescension they have since received. But the New Atheist still ought to feel betrayed. Why isn't this an equally correct criticism of *everything else*?

While the atheists were going around saying there was no God, the environmentalists were going around saying climate change was real. The feminists were going around saying sexism was bad. And the Democrats were going around saying Donald Trump was an awful person. All of these statements might be controversial *somewhere*, but meet basically zero resistance in educated urban liberal spaces. All get repeated day-in and day-out by groups of people who make entire careers out of repeating them. And all get said in the same condescending way, a sort of society-wide plague of Voxsplaining.

This is 90% of popular intellectual culture these days: progressives regurgitating progressivism to other progressives for nothing but the warm glow of being told "Yup, that was some good progressivizing there". Conservatives make fun of this incessantly, and they are right to do so. But for some reason, in the case of New Atheism and *only* in the case of New Atheism, Progressivism itself suddenly turned and said "Hey, you're just repeating our own platitudes back to us!" And New Atheism, caught flat-footed, mouth open wide: "But... but... we thought we were supposed to... we thought..." .

Think of one of those corrupt kleptocracies where the dictator takes bribes, all his ministers take bribes, all their assistants take bribes, the anti-corruption task force takes bribes, etc. Then one day some shmuck manages to get on the dictator's bad side and – bam – the secret police nab him for taking bribes. The look on his face the moment before the firing squad shoots – that's how I imagine New Atheists feeling too.

So who's the dictator in this analogy? And what did New Atheism do to get on their bad side?

Maybe New Atheism failed to make the case that it was socially important. All these movements have a mix of factual claims and social calls to action – climate change activism combines "we should accept the scientifically true fact that the climate is changing" with "we should worry about climate change causing famines, hurricanes, etc", just as atheism combines "we should accept the scientifically true fact that God does not exist" with "we should worry about religion's promotion of terrorism, homophobia, et cetera". But the climate change people seem better at sounding like they care about the people involved, compared to atheists usually sounding more concerned with Truth For Its Own Sake and

bringing in the other stuff as a justification.

Or maybe the New Atheists just didn't know how to stay relevant. Trump resistance always has new tweets to keep its attention. Social justice always has a new sexist celebrity to be angry about. Sure, a few New Atheists tried to keep up with the latest secretly-gay televangelist, but most of them kept going about intricacies of the kalam argument that had been done to death by 1400 AD. This is just an example – maybe there are other asymmetries that are more important?

Maybe the New Atheists accidentally got on board just before a nascent [Grey Tribe/Blue Tribe split](#) and tried to get Blue Tribe credibility by sending Grey Tribe signals. At some point there was a cultural fissure between Acela Corridor thinkfluencers with humanities degrees and Silicon Valley bloggers with STEM degrees, and the former got a head start on hating the latter while the latter still thought everybody was on the same anti-Republican side.

And the cynic in me wonders whether New Atheism wasn't pointless and obvious *enough*. There are more church-goers in educated liberal circles than Trump supporters, climate deniers, or self-identified racists. Maybe that made the "repeat platitudes to people who already believe them" game a little less fun, caused some friction – "You're talking about my dear grandmother!"

I don't know. The whole problem is so strange. For a brief second, modern culture looked at New Atheism, saw itself, and said "Huh, this is really stupid and annoying". Then it cast New Atheism into the outer darkness while totally failing to generalize that experience to anything else. Why would it do that? Could it happen again? *Please* can it happen again? *Pretty* please?

The Influenza Of Evil

Posted on January 13, 2015 by Scott Alexander



I

A recent Cracked piece: [Five Everyday Groups Society Says It's Okay To Mock](#). It begins:

There's a rule in comedy that says you shouldn't punch down. It's okay to make fun of someone rich and famous, because they're too busy molesting groupies with 100-dollar bills to notice, but if you make a joke at the expense of a homeless person, you're just an asshole. That said, we as a society have somehow decided on a few arbitrary exceptions to this rule.

"Somehow decided on a few arbitrary exceptions" isn't very technical. Let's see if we can do better.

Earlier this week, I wrote about [things that are anti-inductive](#). Something is anti-inductive if it fights back against your attempts to understand it. The classic example is the stock market. If someone learns that the stock market is always low on Tuesdays, then they'll buy lots of stocks on Tuesdays to profit from the anomaly. But this raises the demand for stocks on Tuesdays, and therefore stocks won't be low on Tuesdays anymore. To detect a pattern is to destroy the pattern.

The less classic example is job interviews where every candidate is trying to distinguish themselves from every other candidate. If someone learns that interviewers are impressed if you talk about your experience in tropical medicine, then as more and more people catch on they'll all get experience in tropical medicine, it will become cliche, and people won't be impressed by it anymore.

Evil, too, is anti-inductive.

The Nazis were very successful evildoers, at least for a while. Part of their success was convincing people – at least the German people, but sometimes also foreigners – that they were the good guys. And they *were* able to convince a lot of people, because people can be pretty dumb, a lot of them kind of just operate by pattern-matching, and the Nazis didn't match enough patterns to set off people's alarms.

Neo-Nazis cannot be called "successful" in any sense of the word. Their PR problem isn't just that they're horrible – a lot of groups are horrible and do much better than neo-Nazis. Their PR problem is that they're horrible in exactly the way that our culture formed memetic antibodies against. Our pattern-matching faculties have been trained on Nazis being evil. The alarm bells that connect everything about Nazis to evil are hypersensitive, so much so that even contingent features of the Nazis remain universally acknowledged evil-signals.



It would be premature to say that we will never have to worry about fascism again. But for now, we are probably pretty safe from fascism that starts its sales pitch with “Hi, I’m fascism! Want a swastika armband?”

Huey Long [supposedly](#) predicted that “Fascism in America will attempt to advance under the banner of anti-fascism.” I’m not sure I like the saying as it stands – it seems too susceptible to Hitler Jr. telling Churchill Jr. that he’s marching under the banner of anti-fascism which proves he’s the *real* fascist. Then again, in a world where capitalism marches under the banner of [“socialism with Chinese characteristics”](#), who knows? I would prefer to say that fascism will, at the very least, advance in a way which carefully takes our opposition to fascism into account.

Sure enough, people who had learned to be wary of fascism were still highly susceptible to communism, which wore its anti-fascism proudly on its sleeve as a symbol of how great it was. It convinced a lot of very smart people in the free

world that it was the best thing since sliced bread, all while murdering tens of millions of people. Meanwhile, our memetic immune systems were sitting watchfully at their posts, thinking “Well, this doesn’t look at all like Nazism. They’re saying all the right stuff about equality, which is like the *opposite* of what the Nazis said. I’m giving them a pass.”

In fact, I’ll make the analogy more explicit. Every winter, there’s a flu epidemic. Every spring and summer, people’s bodies put in a lot of effort making antibodies to *last year’s* flu. The next winter, the flu mutates a little, a new virus with new antigens starts a new epidemic, and the immune system doesn’t have a clue: “This virus doesn’t have the very very specific characteristic I’ve learned to associate with the flu. Maybe it wants to be my friend!” This is why we need the WHO to predict what the up-and-coming flu virus will be and give us vaccines against it; it’s also why their job is so hard; they don’t know what’s coming, except that it will look different from however it’s looked before.

Nowadays most people’s memetic immune systems have some antibodies to communism, and people talking with Russian accents about how we need to eliminate the bourgeoisie and institute a dictatorship of the proletariat sends shiver up the spines of a lot of people. Nowadays an openly Communist party faces the same uphill battle as an openly Nazi party.

But that just means that if there’s some other evil on the horizon, it probably won’t resemble either fascism or communism. It will be movement about which everyone’s saying “These new guys are so great! They don’t pattern-match to any of the kinds of evil we know about at all!” By Long’s formulation, it may very well be marching under the banners of anti-fascism and anti-Communism.

(I’m not vagueblogging, by the way. I honestly don’t have anyone in mind here. The whole point is that it’s probably someone I’m not expecting. And if you say “I KNOW EXACTLY WHICH GROUP IT WILL BE, BASED ON THOSE CRITERIA IT’S CLEARLY X!” consider the possibility that you’re missing the point.)

II

But getting back to the Cracked article.

We as a society have mostly figured out that shouting “GET A JOB, LOSER!” at the homeless is mean. We have mostly figured out that shouting “YOU’RE GOING TO HELL” at people of different religions is bad. We’re even, slowly but surely, starting to wonder whether there’s something problematic about shouting “FAGGOTS!” at the local gay couple.

Stupid bullies will continue to do those things, just as stupid investors will continue to read “How To Beat The Stock Market” books published in 1985, and stupid socialites will continue to wear the fashion that was cool six months ago.

But smart bullies are driven by their desire to have their bullying make them more popular, to get the rest of the world pointing and laughing with them. In a [Blue Tribe](#) bubble, shouting “FAGGOT” at gay people is no longer a good way to do that. The smart bullies in these circles have long since stopped shouting at gays – not because they’ve become any nicer, but because that’s no longer the best way to keep their audience laughing along with them.

Cracked starts off by naming mentally ill celebrities as a group society considers it okay to mock. This doesn’t seem surprising. Nowadays people talk a lot about punching-up versus punching-down. But that just means bullies who want to successfully punch down will come up with a way to make it look like they’re punching up. Take a group that’s high-status and wealthy, but find a subset who are actually in serious trouble and mock *them*, all the while shouting “I’M PUNCHING UP, I’M PUNCHING UP!”. Thus mentally ill celebrities.

The other examples are harder to figure out. I would argue that they’re ones that are easy to victim-blame (ie obesity),

ones that punch down on axes orthogonal to the rich-poor axis we usually think about and so don't look like punching down (ie virginity), or ones that are covertly associated with an outgroup. In every case, I would expect the bullies involved, when they're called upon, it to loudly protest "But that's not real bullying! It's not like [much more classic example of bullying, like mocking the homeless]!" And they will be right. It's just different enough to be the hot new bullying frontier that most people haven't caught onto yet.

I think the Cracked article is doing good work. It's work that I also try to do (see for example [number 6 here](#), which corresponds to Cracked's number 5). It's the work of pointing these things out, saying "Actually, no, that's bullying", until eventually it sinks into the culture, the bullies realize they'll be called out if they keep it up, and they move on to some new target.

All of this ties *way* into the dynamic I talked about in *Untitled*. I mean, look at the people on Cracked's list of whom society says it's okay to mock. Virgins. The obese. People who live in their parents' basements. Generalize "mentally ill celebrities" just a little bit to get "people who are financially well-off but non-neurotypical" and there you go.

I apologize for irresponsibly claiming to have found a pattern in an anti-inductive domain. You may now all adjust your behavior to make me wrong.

Guided By The Beauty Of Our Weapons

Posted on March 24, 2017 by Scott Alexander



Content note: kind of talking around Trump supporters and similar groups as if they're not there.

I

Tim Harford writes [The Problem With Facts](#), which uses Brexit and Trump as jumping-off points to argue that people are mostly impervious to facts and resistant to logic:

All this adds up to a depressing picture for those of us who aren't ready to live in a post-truth world. Facts, it seems, are toothless. Trying to refute a bold, memorable lie with a fiddly set of facts can often serve to reinforce the myth. Important truths are often stale and dull, and it is easy to manufacture new, more engaging claims. And giving people more facts can backfire, as those facts provoke a defensive reaction in someone who badly wants to stick to their existing world view. "This is dark stuff," says Reifler. "We're in a pretty scary and dark time."

He admits he has no easy answers, but cites some studies showing that "scientific curiosity" seems to help people become interested in facts again. He thinks maybe we can inspire scientific curiosity by linking scientific truths to human interest stories, by weaving compelling narratives, and by finding "a Carl Sagan or David Attenborough of social science".

I think this is generally a good article and makes important points, but there are three issues I want to highlight as possibly pointing to a deeper pattern.

First, the article makes the very strong claim that "facts are toothless" – then tries to convince its readers of this using facts. For example, the article highlights a study by Nyhan & Reifler which finds a "backfire effect" – correcting people's misconceptions only makes them cling to those misconceptions more strongly. Harford expects us to be impressed by this study. But how is this different from all of those social science facts to which he believes humans are mostly impervious?

Second, Nyhan & Reifler's work on the backfire effect is probably not true. The original study establishing its existence [failed](#) to replicate (see eg [Porter & Wood, 2016](#)). This isn't directly contrary to Harford's argument, because Harford doesn't cite the original study – he cites a slight extension of it done a year later by the same team that comes to a slightly different conclusion. But given that the entire field is now in serious doubt, I feel like it would have been judicious to mention some of this in the article. This is especially true given that the article itself is about the way that false ideas spread by people never double-checking their beliefs. It seems to me that if you believe in an epidemic of falsehood so widespread that the very ability to separate fact from fiction is under threat, it ought to inspire a state of [CONSTANT VIGILANCE](#), where you obsessively question each of your beliefs. Yet Harford writes an entire article about a worldwide plague of false beliefs without mustering enough vigilance to see if the relevant studies are true or not.

Third, Harford describes his article as being about *agnostology*, "the study of how ignorance is deliberately produced". His key example is tobacco companies sowing doubt about the negative health effects of smoking – for example, he

talks about tobacco companies sponsoring (accurate) research into all of the non-smoking-related causes of disease so that everyone focused on those instead. But his solution – telling engaging stories, adding a human interest element, enjoyable documentaries in the style of Carl Sagan – seems unusually unsuited to the problem. The National Institute of Health can make an engaging human interest documentary about a smoker who got lung cancer. And the tobacco companies can make an engaging human interest documentary about a guy who got cancer because of asbestos, then was saved by tobacco-sponsored research. Opponents of Brexit can make an engaging documentary about all the reasons Brexit would be bad, and then proponents of Brexit can make an engaging documentary about all the reasons Brexit would be good. If you get good documentary-makers, I assume both will be equally convincing regardless of what the true facts are.

All three of these points are slightly unfair. The first because Harford's stronger statements about facts are probably exaggerations, and he just meant that in *certain cases* people ignore evidence. The second because the specific study cited wasn't the one that failed to replicate and Harford's thesis might be that it was different enough from the original that it's probably true. And the third because the documentaries were just one idea meant to serve a broader goal of increasing "scientific curiosity", a construct which has been shown in studies to be helpful in getting people to believe true things.

But I worry that taken together, they suggest an unspoken premise of the piece. It isn't that *people* are impervious to facts. Harford doesn't expect his reader to be impervious to facts, he doesn't expect documentary-makers to be impervious to facts, and he certainly doesn't expect *himself* to be impervious to facts. The problem is that there's some weird tribe of fact-immune troglodytes out there, going around refusing vaccines and voting for Brexit, and the rest of us have to figure out what to do about them. The fundamental problem is one of *transmission*: how can we make knowledge percolate down from the fact-loving elite to the fact-impervious masses?

And I don't want to condemn this too hard, because it's obviously true up to a point. Medical researchers have lots of useful facts about vaccines. Statisticians know some great facts about the link between tobacco and cancer (shame about [Ronald Fisher](#), though). Probably there are even some social scientists who have a fact or two.

Yet [as I've argued before](#), excessive focus on things like vaccine denialists teaches the wrong habits. It's a desire to take a degenerate case, the rare situation where one side is obviously right and the other bizarrely wrong, and make it into the flagship example for modeling all human disagreement. Imagine a theory of jurisprudence designed only to smack down sovereign citizens, or a government pro-innovation policy based entirely on warning inventors against perpetual motion machines.

And in this wider context, part of me wonders if the focus on transmission is part of the problem. Everyone from statisticians to Brexiteers knows that they are right. The only remaining problem is how to convince others. Go on Facebook and you will find a million people with a million different opinions, each confident in her own judgment, each zealously devoted to informing everyone else.

Imagine a classroom where everyone believes they're the teacher and everyone else is students. They all fight each other for space at the blackboard, give lectures that nobody listens to, assign homework that nobody does. When everyone gets abysmal test scores, one of the teachers has an idea: *I need a more engaging curriculum*. Sure. That'll help.

II

A new Nathan Robinson article: [Debate Vs. Persuasion](#). It goes through the same steps as the Harford article, this time from the perspective of the political Left. Deploying what Robinson calls "Purely Logical Debate" against Trump supporters hasn't worked. Some leftists think the answer is violence. But this may be premature; instead, we should

try the tools of rhetoric, emotional appeal, and other forms of discourse that aren't Purely Logical Debate. In conclusion, Bernie Would Have Won.

I think giving up on argumentation, reason, and language, just because Purely Logical Debate doesn't work, is a mistake. It's easy to think that if we can't convince the right with facts, there's no hope at all for public discourse. But this might not suggest anything about the possibilities of persuasion and dialogue. Instead, it might suggest that mere facts are rhetorically insufficient to get people excited about your political program.

The resemblance to Harford is obvious. You can't convince people with facts. But you *might* be able to convince people with facts carefully intermixed with human interest, compelling narrative, and emotional appeal.

Once again, I think this is generally a good article and makes important points. But I still want to challenge whether things are quite as bad as it says.

Google "[debating Trump supporters is](#)", and you realize where the article is coming from. It's page after page of "debating Trump supporters is pointless", "debating Trump supporters is a waste of time", and "debating Trump supporters is like [funny metaphor for thing that doesn't work]". The overall picture you get is of a world full of Trump opponents and supporters debating on every street corner, until finally, after months of banging their heads against the wall, everyone collectively decided it was futile.

Yet I have the opposite impression. Somehow a sharply polarized country went through a historically divisive election with *essentially no debate taking place*.

Am I about to [No True Scotsman](#) the hell out of the word "debate"? Maybe. But I feel like in using the exaggerated phrase "Purely Logical Debate, Robinson has given me leave to define the term as strictly as I like. So here's what I think are minimum standards to deserve the capital letters:

1. Debate where two people with opposing views are *talking* to each other (or writing, or IMing, or some form of bilateral communication). Not a pundit putting an article on *Huffington Post* and demanding Trump supporters read it. Not even a Trump supporter who comments on the article with a counterargument that the author will never read. Two people who have chosen to engage and to listen to one another.
2. Debate where both people want to be there, and have chosen to enter into the debate in the hopes of getting something productive out of it. So not something where someone posts a "HILLARY IS A CROOK" meme on Facebook, someone gets really angry and lists all the reasons Trump is an even bigger crook, and then the original poster gets angry and has to tell them why they're wrong. Two people who have made it their business to come together at a certain time in order to compare opinions.
3. Debate conducted in the spirit of mutual respect and collaborative truth-seeking. Both people reject personal attacks or 'gotcha' style digs. Both people understand that the other person is *around* the same level of intelligence as they are and may have some useful things to say. Both people understand that they themselves might have some false beliefs that the other person will be able to correct for them. Both people go into the debate with the hope of convincing their opponent, but not completely rejecting the possibility that their opponent might convince them also.
4. Debate conducted outside of a high-pressure point-scoring environment. No audience cheering on both participants to respond as quickly and bitingly as possible. If it can't be done online, at least do it with a smartphone around so you can open Wikipedia to resolve simple matters of fact.
5. Debate where both people agree on what's being debated and try to stick to the subject at hand. None of this

"I'm going to vote Trump because I think Clinton is corrupt" followed by "Yeah, but Reagan was even worse and that just proves you Republicans are hypocrites" followed by "We're hypocrites? You Democrats claim to support women's rights but you love Muslims who make women wear headscarves!" Whether or not it's hypocritical to "support women's rights" but "love Muslims", it doesn't seem like anyone is even *trying* to change each other's mind about Clinton at this point.

These to me seem like the *bare minimum* conditions for a debate that could possibly be productive.

(and while I'm asking for a pony on a silver platter, how about both people have to read [How To Actually Change Your Mind](#) first?)

Meanwhile, in reality...

If you search "debating Trump supporters" without the "is", your first result is [this video](#), where some people with a microphone corner some other people at what looks like a rally. I can't really follow the conversation because they're all shouting at the same time, but I can make out somebody saying 'Republicans give more to charity!' and someone else responding 'That's cause they don't do anything at their jobs!'. Okay.

The second link is [this podcast](#) where a guy talks about debating Trump supporters. After the usual preface about how stupid they were, he describes a typical exchange – "It's kind of amazing how they want to go back to the good old days... Well, when I start asking them 'You mean the good old days when 30% of the population were in unions'... they never seem to like to hear that!... so all this unfettered free market capitalism has got to go bye-bye. They don't find comfort in that idea either. It's amazing. I can say I now know what cognitive dissonance feels like on someone's face." I'm glad time travel seems to be impossible, because otherwise I would be tempted to warp back and change my vote to Trump just to spite this person.

The third link is Vanity Fair's ["Foolproof Guide To Arguing With Trump Supporters"](#), which suggests "using their patriotism against them" by telling them that wanting to "curtail the rights and privileges of certain of our citizens" is un-American.

I worry that people do this kind of thing every so often. Then, when it fails, they conclude "Trump supporters are immune to logic". This is much like observing that Republicans go out in the rain without melting, and concluding "Trump supporters are immortal".

Am I saying that if you met with a conservative friend for an hour in a quiet cafe to talk over your disagreements, they'd come away convinced? No. I've changed my mind on various things during my life, and it was never a single moment that did it. It was more of a series of different things, each taking me a fraction of the way. As the old saying goes, "First they ignore you, then they laugh at you, then they fight you, then they fight you half-heartedly, then they're neutral, then they then they grudgingly say you might have a point even though you're annoying, then they say on balance you're mostly right although you ignore some of the most important facets of the issue, then you win."

There might be a parallel here with the one place I see something like Purely Logical Debate on a routine basis: cognitive psychotherapy. I know this comparison sounds crazy, because psychotherapy is supposed to be the opposite of a debate, and trying to argue someone out of their delusions or depression inevitably fails. The rookiest of all rookie therapist mistakes is to say "FACT CHECK: The patient says she is a loser who everybody hates. PsychiaFact rates this claim: PANTS ON FIRE."

But in other ways it's a lot like the five points above. You have two people who disagree – the patient thinks she's a worthless loser who everyone hates, and the therapist thinks maybe not. They meet together in a spirit of voluntary mutual inquiry, guaranteed safe from personal attacks like "You're crazy!". Both sides go over the evidence together,

sometimes even agreeing on explicit experiments like “Ask your boyfriend tonight whether he hates you or not, predict beforehand what you think he’s going to say, and see if your prediction is accurate”. And both sides approach the whole process suspecting that they’re right but admitting the possibility that they’re wrong (very occasionally, after weeks of therapy, I realize that frick, everyone really *does* hate my patient. Then we switch strategies to helping her with social skills, or helping her find better friends).

And contrary to what you see in movies, this doesn’t usually give a single moment of blinding revelation. If you spent your entire life talking yourself into the belief that you’re a loser and everyone hates you, no single fact or person is going to talk you out of it. But after however many months of intensive therapy, sometimes someone who was *sure* that they were a loser is now *sort of questioning* whether they’re a loser, and has the mental toolbox to take things the rest of the way themselves.

This was also the response I got when I tried to make [an anti-Trump case](#) on this blog. I don’t think there were any sudden conversions, but here were some of the positive comments I got from Trump supporters:

- “This is a compelling case, but I’m still torn.”
- “This contains the most convincing arguments for a Clinton presidency I have ever seen. But, perhaps also unsurprisingly, while it did manage to shift some of my views, it did not succeed in convincing me to change my bottom line.”
- “This article is perhaps the best argument I have seen yet for Hillary. I found myself nodding along with many of the arguments, after this morning swearing that there was nothing that could make me consider voting for Hillary... the problem in the end was that it wasn’t enough.”
- “The first coherent article I’ve read justifying voting for Clinton. I don’t agree with your analysis of the dollar “value” of a vote, but other than that, something to think about.”
- “Well I don’t like Clinton at all, and I found this essay reasonable enough. The argument from continuity is probably the best one for voting Clinton if you don’t particularly love any of her policies or her as a person. Trump is a wild card, I must admit.”
- As an orthodox Catholic, you would probably classify me as part of your conservative audience... I certainly concur with both the variance arguments and that he’s not conservative by policy, life, or temperament, and I will remain open to hearing what you have to say on the topic through November.
- “I’ve only come around to the ‘hold your nose and vote Trump’ camp the past month or so... I won’t say [you] didn’t make me squirm, but I’m holding fast to my decision.”

These are the people you say are completely impervious to logic so don’t even try? It seems to me like this argument was one of not-so-many straws that might have broken some camels’ backs if they’d been allowed to accumulate. And the weird thing is, when I re-read the essay I notice a lot of flaws and things I wish I’d said differently. I don’t think it was an exceptionally good argument. I think it was... an argument. It was something more than saying “You think the old days were so great, but the old days had labor unions, CHECKMATE ATHEISTS”. This isn’t what you get when you do a splendid virtuous performance. This is what you get *when you show up*.

(and lest I end up ‘objectifying’ Trump supporters as prizes to be won, I’ll add that in the comments some people made pro-Trump arguments, and two people who were previously leaning Clinton said that they were feeling uncomfortably close to being convinced)

Another SSC story. I keep trying to keep “culture war”-style political arguments from overrunning the blog and subreddit, and every time I add restrictions [a bunch of people complain](#) that this is the only place they can go for that.

Think about this for a second. A heavily polarized country of three hundred million people, split pretty evenly into two sides and obsessed with politics, blessed with the strongest free speech laws in the world, and people are complaining that I can't change my comment policy because this one small blog is *the only place they know where they can debate people from the other side*.

Given all of this, I reject the argument that Purely Logical Debate has been tried and found wanting. Like GK Chesterton, I think it has been found difficult and left untried.

III

Therapy might change minds, and so might friendly debate among equals, but neither of them scales very well. Is there anything that big fish in the media can do beyond the transmission they're already trying?

Let's go back to that Nyhan & Reifler study which found that fact-checking backfired. As I mentioned above, a replication attempt by Porter & Wood found the opposite. This could have been the setup for a nasty conflict, with both groups trying to convince academia and the public that they were right, or even accusing the other of scientific malpractice.

Instead, something great happened. All four researchers [decided to work together](#) on an "adversarial collaboration" – a bigger, better study where they all had input into the methodology and they all checked the results independently. The collaboration found that fact-checking generally didn't backfire in most cases. All four of them used their scientific clout to publicize the new result and launch further investigations into the role of different contexts and situations.

Instead of treating disagreement as demonstrating a need to transmit their own opinion more effectively, they viewed it as demonstrating a need to collaborate to investigate the question together.

And yeah, part of it was that they were all decent scientists who respected each other. But they didn't *have* to be. If one team had been total morons, and the other team was secretly laughing at them the whole time, the collaboration still would have worked. All required was an assumption of good faith.

A while ago I blogged about a journalistic spat between German Lopez and Robert VerBruggen on gun control. Lopez wrote [a voxsplainer](#) citing some statistics about guns. VerBruggen wrote [a piece at National Review](#) saying that some of the statistics were flawed. German fired back (pun not intended) [with an article](#) claiming that VerBruggen was ignoring better studies.

(Then I [yelled at both of them](#), as usual.)

Overall the exchange was in the top 1% of online social science journalism – by which I mean it included at least one statistic and at some point that statistic was superficially examined. But in the end, it was still just two people arguing with one another, each trying to transmit his superior knowledge to each other and the reading public. As good as it was, it didn't meet my five standards above – and nobody expected it to.

But now I'm thinking – what would have happened if Lopez and VerBruggen had joined together in an adversarial collaboration? Agreed to work together to write an article on gun statistics, with nothing going into the article unless they both approved, and then they both published that article on their respective sites?

This seems like a mass media equivalent of shifting from Twitter spats to serious debate, from transmission mindset to collaborative truth-seeking mindset. The adversarial collaboration model is just the first one to come to mind right now. I've blogged about others before – for example, bets, prediction markets, and calibration training.

The media already spends a lot of effort *recommending* good behavior. What if they tried *modeling* it?

IV

The bigger question hanging over all of this: “Do we *have* to?”

Harford’s solution – compelling narratives and documentaries – sounds easy and fun. Robinson’s solution – rhetoric and emotional appeals – also sounds easy and fun. Even the solution Robinson rejects – violence – is easy, and fun for a certain type of person. All three work on pretty much anybody.

Purely Logical Debate is difficult and annoying. It doesn’t scale. It only works on the subset of people who are willing to talk to you in good faith and smart enough to understand the issues involved. And even then, it only works glacially slowly, and you win only partial victories. What’s the point?

Logical debate has one advantage over narrative, rhetoric, and violence: it’s an *asymmetric weapon*. That is, it’s a weapon which is stronger in the hands of the good guys than in the hands of the bad guys. In ideal conditions (which may or may not ever happen in real life) – the kind of conditions where everyone is charitable and intelligent and wise – the good guys will be able to present stronger evidence, cite more experts, and invoke more compelling moral principles. The whole point of logic is that, when done right, it can only prove things that are true.

Violence is a *symmetric weapon*; the bad guys’ punches hit just as hard as the good guys’ do. It’s true that hopefully the good guys will be more popular than the bad guys, and so able to gather more soldiers. But this doesn’t mean violence itself is asymmetric – the good guys will only be more popular than the bad guys insofar as their ideas have previously spread through some means other than violence. Right now antifascists outnumber fascists and so could probably beat them in a fight, but antifascists didn’t come to outnumber fascists by winning some kind of primordial fistfight between the two sides. They came to outnumber fascists because people rejected fascism on the merits. These merits might not have been “logical” in the sense of Aristotle dispassionately proving lemmas at a chalkboard, but “fascists kill people, killing people is wrong, therefore fascism is wrong” is a sort of folk logical conclusion which is both correct and compelling. Even “a fascist killed my brother, so fuck them” is a placeholder for a powerful philosophical argument making a probabilistic generalization from indexical evidence to global utility. So insofar as violence is asymmetric, it’s because it parasitizes on logic which allows the good guys to be more convincing and so field a bigger army. Violence itself doesn’t enhance that asymmetry; if anything, it decreases it by giving an advantage to whoever is more ruthless and power-hungry.

The same is true of documentaries. As I said before, Harford can produce as many anti-Trump documentaries as he wants, but Trump can fund documentaries of his own. He has the best documentaries. Nobody has ever seen documentaries like this. They’ll be absolutely huge.

And the same is true of rhetoric. Martin Luther King was able to make persuasive emotional appeals for good things. But Hitler was able to make persuasive emotional appeals for bad things. I’ve [previously argued](#) that Mohammed counts as the most successful persuader of all time. These three people pushed three very different ideologies, and rhetoric worked for them all. Robinson writes as if “use rhetoric and emotional appeals” is a novel idea for Democrats, but it seems to me like they were doing little else throughout the election (pieces attacking Trump’s character, pieces talking about how inspirational Hillary was, pieces appealing to various American principles like equality, et cetera). It’s just that they did a bad job, and Trump did a better one. The real takeaway here is “do rhetoric better than the other guy”. But “succeed” is not a primitive action.

Unless you use asymmetric weapons, the best you can hope for is to win by coincidence.

That is, there's no reason to think that good guys are consistently better at rhetoric than bad guys. Some days the Left will have an Obama and win the rhetoric war. Other days the Right will have a Reagan and *they'll* win the rhetoric war. Overall you should average out to a 50% success rate. When you win, it'll be because you got lucky.

And there's no reason to think that good guys are consistently better at documentaries than bad guys. Some days the NIH will spin a compelling narrative and people will smoke less. Other days the tobacco companies will spin a compelling narrative and people will smoke more. Overall smoking will stay the same. And again, if you win, it's because you lucked out into having better videographers or something.

I'm not against winning by coincidence. If I stumbled across Stalin and I happened to have a gun, I would shoot him without worrying about how it's "only by coincidence" that he didn't have the gun instead of me. You should use your symmetric weapons if for no reason other than that the other side's going to use *theirs* and so you'll have a disadvantage if you don't. But you shouldn't confuse it with a long-term solution.

Improving the quality of debate, shifting people's mindsets from transmission to collaborative truth-seeking, is a painful process. It has to be done one person at a time, it only works on people who are already *almost* ready for it, and you will pick up far fewer warm bodies per hour of work than with any of the other methods. But in an otherwise-random world, even a little purposeful action can make a difference. Convincing 2% of people would have flipped three of the last four US presidential elections. And this is a capacity to win-for-reasons-other-than-coincidence that you can't build any other way.

(and my hope is that the people most willing to engage in debate, and the ones most likely to recognize truth when they see it, are disproportionately influential – scientists, writers, and community leaders who have influence beyond their number and can help others see reason in turn)

I worry that I'm not communicating how beautiful and inevitable all of this is. We're surrounded by a vast confusion, "a darkling plain where ignorant armies clash by night", with one side or another making a temporary advance and then falling back in turn. And in the middle of all of it, there's this gradual capacity-building going on, where what starts off as a hopelessly weak signal gradually builds up strength, until one army starts winning a little more often than chance, then a lot more often, and finally takes the field entirely. Which seems strange, because surely you can't build any complex signal-detection machinery in the middle of all the chaos, surely you'd be shot the moment you left the trenches, but – *your enemies are helping you do it*. Both sides are diverting their artillery from the relevant areas, pooling their resources, helping bring supplies to the engineers, because until the very end they think it's going to ensure *their* final victory and not yours.

You're doing it right under their noses. They might try to ban your documentaries, heckle your speeches, fight your violence Middlebury-student-for-Middlebury-student – but when it comes to the long-term solution to ensure your complete victory, they'll roll down their sleeves, get out their hammers, and build it alongside you.

A parable: Sally is a psychiatrist. Her patient has a strange delusion: that *Sally* is the patient and *he* is the psychiatrist. She would like to commit him and force medication on him, but he is an important politician and if push comes to shove he might be able to commit *her* instead. In desperation, she proposes a bargain: they will *both* take a certain medication. He agrees; from within his delusion, it's the best way for him-the-psychiatrist to cure her-the-patient. The two take their pills at the same time. The medication works, and the patient makes a full recovery.

(well, half the time. The other half, the medication works and *Sally* makes a full recovery.)

V

Harford's article says that facts and logic don't work on people. The various lefty articles say they merely don't work on Trump supporters, ie 50% of the population.

If you genuinely believe that facts and logic don't work on people, you shouldn't be writing articles with potential solutions. You should be jettisoning everything you believe and entering a state of pure Cartesian doubt, where you try to rederive everything from *cogito ergo sum*.

If you genuinely believe that facts and logic don't work on at least 50% of the population, again, you shouldn't be writing articles with potential solutions. You should be worrying whether you're in that 50%. After all, how did you figure out you aren't? By using facts and logic? *What did we just say?* Nobody is doing either of these things, so I conclude that they accept that facts can sometimes work. Asymmetric weapons are not a pipe dream. As Gandhi used to say, "If you think the world is all bad, remember that it contains people like you."

You are not completely immune to facts and logic. But you have been wrong about things before. You may be a bit smarter than the people on the other side. You may even be a *lot* smarter. But fundamentally their problems are your problems, and the same kind of logic that convinced you can convince them. It's just going to be a long slog. You didn't develop *your* opinions after a five-minute shouting match. You developed them after years of education and acculturation and engaging with hundreds of books and hundreds of people. Why should they be any different?

You end up believing that the problem is deeper than insufficient documentary production. The problem is that Truth is a weak signal. You're trying to perceive Truth. You would like to hope that the other side is trying to perceive Truth too. But at least one of you is doing it wrong. It seems like perceiving Truth accurately is harder than you thought.

You believe your mind is a truth-sensing instrument that does at least a little bit better than chance. You *have* to believe that, or else what's the point? But it's like one of those physics experiments set up to detect gravitational waves or something, where it has to be in a cavern five hundred feet underground in a lead-shielded chamber atop a gyroscopically stable platform cooled to one degree above absolute zero, trying to detect fluctuations of a millionth of a centimeter. Except you don't have the cavern or the lead or the gyroscope or the coolants. You're on top of an erupting volcano being pelted by meteorites in the middle of a hurricane.

If you study psychology for ten years, you can remove the volcano. If you spend another ten years obsessively checking your performance in various *metis*-intensive domains, you can remove the meteorites. You can never remove the hurricane and you shouldn't try. But if there are a thousand trustworthy people at a thousand different parts of the hurricane, then the stray gusts of wind will cancel out and they can average their readings to get something approaching a signal.

All of this is too slow and uncertain for a world that needs more wisdom *now*. It would be nice to force the matter, to pelt people with speeches and documentaries until they come around. This will work in the short term. In the long term, it will leave you back where you started.

If you want people to be right more often than chance, you have to teach them ways to distinguish truth from falsehood. If this is in the face of enemy action, you will have to teach them so well that they cannot be fooled. You will have to do it person by person until the signal is strong and clear. You will have to raise the sanity waterline. There is no shortcut.

Ethics Offsets

Posted on January 4, 2015 by Scott Alexander



I

Some people buy voluntary carbon offsets. Suppose they worry about global warming and would feel bad taking a long unnecessary plane trip that pollutes the atmosphere. So instead of not doing it, they take the plane trip, then pay for some environmental organization to clean up an amount of carbon equal to or greater than the amount of carbon they emitted. They're happy because they got their trip, future generations are happy because the atmosphere is cleaner, everyone wins.

We can generalize this to ethics offsets. Suppose you really want to visit an oppressive dictatorial country so you can see the [beautiful tourist sights there](#). But you worry that by going there and spending money, you're propping up the dictatorship. So you take your trip, but you also donate some money to opposition groups and humanitarian groups opposing the dictatorship and helping its victims, at an amount such that you are confident that the oppressed people of the country would prefer you take both actions (visit + donate) than that you take neither action.

I know I didn't come up with this concept, but I'm having trouble finding out who did, so no link for now.

A recent post, [Nobody Is Perfect, Everything Is Commensurable](#), suggests that if you are averse to activism but still feel you have an obligation to improve the world, you can discharge that obligation by giving to charity. This is not quite an ethics offset – it's not exchanging a transgression for a donation so much as saying that a donation is a better way of helping than the thing you were worried about transgressing against anyway – but it's certainly pretty similar.

As far as I can tell, the simplest cases here are 100% legit. I can't imagine anyone saying "You may not take that plane flight you want, even if you donate so much to the environment that in the end it cleans up twice as much carbon dioxide as you produced. You must sit around at home, feeling bored and lonely, and letting the atmosphere be more polluted than if you had made your donation".

But here are two cases I am less certain about.

II

Suppose you feel some obligation to be a vegetarian – either because you believe animal suffering is bad, or you have enough moral uncertainty around the topic for the ethical calculus to come out against. Is it acceptable to continue eating animals, but also donate money to animal rights charities?

A simple example: you eat meat, but also donate money to a group lobbying for cage-free eggs. You are confident that if chickens could think and vote, the average chicken would prefer a world in which you did both these things to a world in which you did neither. This seems to me much like the cases above.

A harder example. You eat meat, but also donate money to a group that convinces people to become vegetarian. Jeff

Kaufman and Brian Tomasik [suggest](#) that about \$10 to \$50 is enough to make one person become vegetarian for one year by sponsoring what are apparently very convincing advertisements.

Eating meat is definitely worth \$1000 per year for me. So if I donate \$1000 to vegetarian advertising, then eat meat, I'm helping turn between twenty and a hundred people vegetarian for a year, and helping twenty to one hundred times as many animals as I would be by becoming vegetarian myself. Clearly this is an excellent deal for me and an excellent deal for animals.

But I still can't help feeling like there's something really wrong here. It's not just the low price of convincing people – even if I was 100% guaranteed that the calculations were right, I'd still feel just as weird. Part of it is a sense of duping others – would they be as eager to become vegetarian if they knew the ads that convinced them were sponsored by meat-eaters?

Maybe! Suppose we go to all of the people convinced by the ads, tell them "I paid for that ad that convinced you, and I still eat meat. Now what?" They answer "Well, I double-checked the facts in the ad and they're all true. That you eat meat doesn't make anything in the advertisement one bit less convincing. So I'm going to stay vegetarian." Now what? Am I off the hook?

A second objection: universalizability. If *everyone* decides to solve animal suffering by throwing money at advertisers, there is no one left to advertise to and nothing gets solved. You just end up with a world where 100% of ads on TVs, in newspapers, and online are about becoming vegetarian, and everyone watches them and says "Well, I'm doing my part! I'm paying for these ads!"

Counter-objection: At that point, no one will be able to say with a straight face that every \$50 spent on ads converts one person to vegetarianism. If I follow the maxim "Either be vegetarian, or donate enough money to be 90% sure I am converting at least two other people to vegetarianism", this maxim *does* universalize, since after animal suffering ads have saturated a certain percent of the population, no one can be 90% sure of convincing anyone else.

As far as I can tell, this is weird but ethical.

III

The second troublesome case is a little more gruesome.

Current estimates suggest that \$3340 worth of donations to global health causes saves, on average, one life.

Let us be excruciatingly cautious and include a two-order-of-magnitude margin of error. At \$334,000, we are *super duper sure* we are saving at least one life.

So. Say I'm a millionaire with a spare \$334,000, and there's a guy I *really* don't like...

Okay, fine. Get the irrelevant objections out of the way first and establish the [least convenient possible world](#). I'm a criminal mastermind, it'll be the perfect crime, and there's zero chance I'll go to jail. I can make it look completely natural, like a heart attack or something, so I'm not going to terrorize the city or waste police time and resources. The guy's not supporting a family and doesn't have any friends who will be heartbroken at his death. There's no political aspect to my grudge, so this isn't going to silence the enemies of the rich or anything like that. I myself have a terminal disease, and so the damage that I inflict upon my own soul with the act – or however it is Leah always phrases it – will perish with me immediately afterwards. There is no God, or if there is one He respects ethics offsets when you get to the Pearly Gates.

Or you know what? Don't get the irrelevant objections out of the way. We can offset those too. The police will waste a lot of time investigating the murder? Maybe I'm *very* rich and I can make a big anonymous donation to the local police force that will more than compensate them for their trouble and allow them to hire extra officers to take up the slack. The local citizens will be scared there's a killer on the loose? They'll forget all about it once they learn taxes have been cut to zero percent thanks to an anonymous donation to the city government from a local tycoon.

Even what seems to me the most desperate and problematic objection – that maybe the malarial Africans saved by global health charities have lives that are in some qualitative way just not as valuable as those of happy First World citizens contributing to the global economy – can be fixed. If I've got enough money, a few hundred thousand to a million ought to be able to save the life of a local person in no way distinguishable from my victim. Heck, since this is a hypothetical problem and I have infinite money, why not save *ten* local people?

The best I can do here is to say that I am crossing a [Schelling fence](#) which might also be crossed by people who will be less scrupulous in making sure their offsets are in order. But perhaps I could offset that too. Also, we could assume I will never tell anybody. Also, anyone can just go murder someone right now without offsetting, so we're not exactly talking about a big temptation for the unscrupulous.

Sort By Controversial

Posted on October 30, 2018 by Scott Alexander



Epistemic status: fiction.

☆☆

Thanks for letting me put my story on your blog. Mainstream media is crap and no one would have believed me anyway.

This starts in September 2017. I was working for a small online ad startup. You know the ads on Facebook and Twitter? We tell companies how to get them the most clicks. This startup – I won’t tell you the name – was going to add deep learning, because investors will throw money at anything that uses the words “deep learning”. We train a network to predict how many upvotes something will get on Reddit. Then we ask it how many likes different ads would get. Then we use whatever ad would get the most likes. [This guy](#) (who is not me) explains it better. Why Reddit? Because the upvotes and downvotes are simpler than all the different Facebook reacts, plus the subreddits allow demographic targeting, plus there’s [an archive of 1.7 billion Reddit comments](#) you can download for training data. We trained a network to predict upvotes of Reddit posts based on their titles.

Any predictive network doubles as a generative network. If you teach a neural net to recognize dogs, you can run it in reverse to get dog pictures. If you train a network to predict Reddit upvotes, you can run it in reverse to generate titles it predicts will be highly upvoted. We tried this and it was pretty funny. I don’t remember the exact wording, but for /r/politics it was something like “Donald Trump is no longer the president. All transgender people are the president.” For r/technology it was about Elon Musk saving Net Neutrality. You can also generate titles that will get maximum downvotes, but this is boring: it will just say things that sound like spam about penis pills.

Reddit has a feature where you can sort posts by controversial. You can see the algorithm [here](#), but tl;dr it multiplies magnitude of total votes (upvotes + downvotes) by balance (upvote:downvote ratio or vice versa, whichever is smaller) to highlight posts that provoke disagreement. Controversy sells, so we trained our network to predict this too. The project went to this new-ish Indian woman with a long name who went by Shiri, and she couldn’t get it to work, so our boss Brad sent me to help. Shiri had tested the network on the big 1.7 billion comment archive, and it had produced controversial-sounding hypothetical scenarios about US politics. So far so good.

The Japanese tested their bioweapons on Chinese prisoners. The Tuskegee Institute tested syphilis on African-Americans. We were either nicer or dumber than they were, because we tested Shiri’s Scissor on ourselves. We had a private internal subreddit where we discussed company business, because Brad wanted all of us to get familiar with the platform. Shiri’s problem was that she’d been testing the controversy-network on our subreddit, and it would just spit out vacuously true or vacuously false statements. No controversy, no room for disagreement. The statement we were looking at that day was about a design choice in our code. I won’t tell you the specifics, but imagine you took every bad and wrong design decision in the world, hard-coded them in the ugliest possible way, and then handed it to the end user with a big middle finger. Shiri’s Scissor spit out, as maximally controversial, the statement that we should design our product that way. We’d spent ten minutes arguing about exactly where the bug was, when Shiri said something about how she didn’t understand why the program was generating obviously true statements.

Shiri’s English wasn’t great, so I thought this was a communication problem. I corrected her. The program was spitting out obviously false statements. She stuck to her guns. I still thought she was confused. I walked her through the

meanings of the English words “true” and “false”. She looked offended. I tried to confirm. She thought this abysmal programming decision, this plan of combining every bad design technique together and making it impossible to ever fix, was the right way to build our codebase? She said it was. Worse, she was confused I *didn’t* think so. She thought this was more or less what we were already doing; it wasn’t. She thought that moving away from this would take a total rewrite and make the code much worse.

At this point I was doubting my sanity, so we went next door to Blake and David, who were senior coders in our company and usually voices of reason. They were talking about their own problem, but I interrupted them and gave them the Scissor statement. Blake gave the reasonable response – why are you bothering me with this stupid wrong garbage? But David had the same confusion Shiri did and started arguing that the idea made total sense. The four of us started fighting. I still was sure Shiri and David just misunderstood the question, even though David was a native English-speaker and the question was crystal-clear. Meanwhile David was feeling more and more condescended to, kept protesting he wasn’t misunderstanding anything, that Blake and I were just crappy programmers who couldn’t make the most basic architecture decisions. He kept insisting the same thing Shiri had, that the Scissor statement had already been the plan and any attempt to go in a different direction would screw everything up. It got so bad that we decided to go to Brad for clarification.

Brad was our founder. Don’t trust the newspapers – not every tech entrepreneur is a greedy antisocial philistine. But everyone in advertising is. Brad definitely was. He was an abrasive amoral son of a bitch. But he was good at charming investors, and he could code, which is more than some bosses. He looked pissed to have the whole coding team come into his office unannounced, but he heard us out.

David tried to explain the issue, but he misrepresented almost every part of it. I couldn’t believe he was lying just to look better to Brad. I cut him off. He told me not to interrupt him. Blake said if he wasn’t lying we wouldn’t have to interrupt to correct him, it degenerated from there. Somehow in the middle of all of this, Brad figured out what we were talking about and he cut us all off. “That’s the stupidest thing I ever heard.” He confirmed it wasn’t the original plan, it was contrary to the original plan, and it was contrary to every rule of good programming and good business. David and Shiri, who were bad losers, accused Blake and me of “poisoning” Brad. David said that of course Brad would side with us. Brad had liked us better from the beginning. We’d racked up cushy project after cushy project while he and Shiri had gotten the dregs. Brad told him he was a moron and should get back to work. He didn’t.

This part of the story ends at 8 PM with Brad firing David and Shiri for a combination of gross incompetence, gross insubordination, and being terrible human beings. With him giving a long speech on how he’d taken a chance on hiring David and Shiri, even though he knew from the beginning that they were unqualified charity cases, and at every turn they’d repaid his kindness with laziness and sabotage. With him calling them a drain on the company and implied they might be working for our competitors. With them calling him an abusive boss, saying the whole company was a scam to trick vulnerable employees into working themselves ragged for Brad’s personal enrichment, and with them accusing us two – me and Blake – of being in on it with Brad.

That was 8 PM. We’d been standing in Brad’s office fighting for five hours. At 8:01, after David and Shiri had stormed out, we all looked at each other and thought – holy shit, the controversial filter works.

I want to repeat that. At no time in our five hours of arguing did this occur to us. We were too focused on the issue at hand, the Scissor statement itself. We didn’t have the perspective to step back and think about how all this controversy came from a statement designed to be maximally controversial. But at 8:01, when the argument was over and we had won, we stepped back and thought – holy shit.

We were too tired to think much about it that evening, but the next day we – Brad and the two remaining members of the coding team – had a meeting. We talked about what we had. Blake gave it its name: Shiri’s Scissor. In some dead language, scissor shares a root with schism. A scissor is a schism-er, a schism-creator. And that was what we had. We

were going to pivot from online advertising to superweapons. We would call the Pentagon. Tell them we had a program that could make people hate each other. Was this ethical? We were in online ads; we would sell our grandmothers to Somali slavers if we thought it would get us clicks. That horse had left the barn a long time ago.

It's hard to just call up the Pentagon and tell them you have a superweapon. Even in Silicon Valley, they don't believe you right away. But Brad called in favors from his friends, and about a week after David and Shiri got fired, we had a colonel from DARPA standing in the meeting room, asking what the hell we thought was so important.

Now we had a problem. We couldn't show the Colonel the Scissor statement that had gotten Dave and Shiri fired. He wasn't in our company; he wasn't even in ad tech; it would seem boring to him. We didn't want to generate a new Scissor statement for the Pentagon. Even Brad could figure out that having the US military descend into civil war would be bad for clicks. Finally we settled on a plan. We explained the concept of Reddit to the Colonel. And then we asked him which community he wanted us to tear apart as a demonstration.

He thought for a second, then said "Mozambique".

We had underestimated the culture gap here. When we asked the Colonel to choose a community to be a Scissor victim, we were expecting "tabletop wargamers" or "My Little Pony fans". But this was not how colonels at DARPA thought about the world. He said "Mozambique". I started explaining to him that this wasn't really how Reddit worked, it needed to be a group with its own subreddit. Brad interrupted me, said that Mozambique had a subreddit.

I could see the wheels turning in Brad's eyes. One wheel was saying "this guy is already skeptical, if we look weak in front of him he'll just write us off completely". The other wheel was calculating how many clicks Mozambique produced. Mene mene tekel upharsin. "Yeah," he said. "Their subreddit is fine. We can do Mozambique."

The Colonel gave us his business card and left. Blake and I were stuck running Shiri's Scissor on the Mozambique subreddit. I know, ethics, but like I said, online ads business, horse, barn door. The only decency we allowed ourselves was to choose the network's tenth pick – we didn't need to destroy everything, just give a demonstration. We got a statement accusing the Prime Minister of disrespecting Islam in a certain way – again, I won't be specific. In the absence of any better method, we PMed the admins of the Mozambique subreddit asking them what they thought. I don't remember what we said, something about being an American political science student learning about Mozambique culture, and could they ask some friends what would happen if the Prime Minister did that specific thing, and then report back to us?

We spent most of a week working on our project to undermine Mozambique. Then we got the news. David and Shiri were suing the company for unfair dismissal and racial discrimination. Brad and Blake and I were white. Shiri was an Indian woman, and David was Jewish. The case should have been laughed out of court – who ever heard of an anti-Semitic Silicon Valley startup? – except that all the documentation showed there was no reason to fire David and Shiri. Their work looked good on paper. They'd always gotten good performance reviews. The company was doing fine – it had even placed ads for more programmers a few weeks before.

David and Shiri knew why they'd been fired. But it didn't matter to them. They were so blinded with hatred for our company, so caught in the grip of the Scissor statement, that they would tell any lie necessary to destroy it. We were caught in a bind. We couldn't admit the existence of Shiri's Scissor, because we were trying to sell it to the Pentagon as a secret weapon, and also, publicly admitting to trying to destroy Mozambique would have been bad PR. But the court was demanding records about what our company had been doing just before and just after the dismissal. A real defense contractor could probably have gotten the Pentagon to write a letter saying our research was classified. But the Pentagon still didn't believe us. The Colonel was humoring us, nothing more. We were stuck.

I don't know how we would have dealt with the legal problems, because what actually happened was Brad went to

David's house and tried to beat him up. You're going to think this was crazy, but you have to understand that David had always been annoying to work with, and that during the argument in Brad's office he had crossed so many lines that, if ever there was a person who deserved physical violence, it was him. Suing the company was just the last straw. I'm not going to judge Brad's actions after he'd spent months cleaning up after David's messes, paying him good money, and then David betrayed him at the end. But anyhow, that was it for our company. Brad got arrested. There was nobody else to pay the bills and keep the lights on. Blake and I were coders and had no idea how to run the business side of things. We handed in our resignations – not literally, Brad was in jail – and that was the end of Name Withheld Online Ad Company, Inc.

We got off easy. That's the takeaway I want to give here. We were unreasonably overwhelmingly lucky. If Shiri and I had started out by arguing about one of the US statements, we could have destroyed the country. If a giant like Google had developed Shiri's Scissor, it would have destroyed Google. If the Scissor statement we generated hadn't just been about a very specific piece of advertising software – if it had been about the tech industry in general, or business in general – we could have destroyed the economy.

As it was, we just destroyed our company and maybe a few of our closest competitors. If you look up internal publications from the online advertising industry around fall 2017, you will find some really weird stuff. [That story about](#) the online ads CEO getting arrested for murder, child abuse, attacking a cop, and three or four other things, and then later it was all found to be [false accusations](#) related to some ill-explained mental disorder – that's the tip of the iceberg. I don't have a good explanation for exactly how the Scissor statement spread or why it didn't spread further, but I bet if I looked into it too much, black helicopters would start hovering over my house. And that's all I'm going to say about that.

As for me, I quit the whole industry. I picked up a job in a more established company using ML for voice recognition, and tried not to think about it too much. I still got angry whenever I thought about the software design issue the Scissor had brought up. Once I saw someone who looked like Shiri at a cafe and I went over intending to give her a piece of my mind. It wasn't her, so I didn't end up in jail with Brad. I checked the news from Mozambique every so often, and it was quiet for a few months, [and then it wasn't](#). I still don't know if we had anything to do with that. Africa just has a lot of conflicts, and if you wait long enough, maybe something will happen. The colonel never tried to get in touch with me. I don't think he ever took us seriously. Maybe he didn't even check the news from Mozambique. Maybe he saw it and figured it was a coincidence. Maybe he tried calling our company, got a message saying the phone was out of service, and didn't think it was worth pursuing. But as time went on and the conflict there didn't get any worse, I hoped the Shiri's Scissor part of my life was drawing to a close.

Then came the Kavanaugh hearings. Something about them gave me a sense of déjà vu. The week of his testimony, I figured it out.

Shiri had told me that when she ran the Scissor on the site in general, she'd just gotten some appropriate controversial US politics scenarios. She had shown me two or three of them as examples. One of them had been very specifically about this situation. A Republican Supreme Court nominee accused of committing sexual assault as a teenager.

This made me freak out. Had somebody gotten hold of the Scissor and started using it on the US? Had that Pentagon colonel been paying more attention than he let on? But why would the Pentagon be trying to divide America? Had some enemy stolen it? I get the *New York Times*, obviously Putin was my first thought here. But how would Putin get Shiri's Scissor? Was I remembering wrong? I couldn't get it out of my head. I hadn't kept the list Shiri had given me, but I had enough of the Scissor codebase to rebuild the program over a few sleepless nights. Then I bought a big blob of compute from Amazon Web Services and threw it at the Reddit comment archive. It took three days and a five-digit sum of money, but I rebuilt the list Shiri must have had. Kavanaugh was in there, just as I remembered. *But so was Colin Kaepernick.*

You've heard of him. He was the football player who refused to stand for the national anthem. If I already knew the Scissor predicted one controversy, why was I so shocked to learn it predicted another? Because Kaepernick started kneeling in 2016. We didn't build the Scissor until 2017. Putin hadn't gotten it from us. Someone had beaten us to it.

Of the Scissor's predicted top hundred most controversial statements, Kavanaugh was #58 and Kaepernick was #42. #86 was the Ground Zero Mosque. #89 was that baker who wouldn't make a cake for a gay wedding. The match isn't perfect, but #99 vaguely looked like the Elian Gonzalez case from 2000. That's five out of a hundred. Is that what would happen by chance? It's a big country, and lots of things happen here, and if a Scissor statement came up in the normal course of events it would get magnified to the national stage. But some of these were too specific. If it was coincidence, I would expect many more near matches than perfect matches. I found only two. The pattern of Scissor statements looked more like someone had arranged them to be perfect fits.

The earliest perfect fit was the Ground Zero Mosque in 2009. Could Putin have had a Scissor-like program in 2009? I say no way. This will sound weird to you if you're not in the industry. Why couldn't a national government have been eight years ahead of an online advertising company? All I can say is: machine learning moves faster than that. Russia couldn't hide a machine learning program that put it eight years ahead of the US. Even the Pentagon couldn't hide a program that put it eight years ahead of industry. The NSA is thirty years ahead of industry in cryptography and everyone knows it.

But then who was generating Scissor statements in 2009? I have no idea. And you know what? I can't bring myself to care.

If you just read a Scissor statement off a list, it's harmless. It just seems like a trivially true or trivially false thing. It doesn't activate until you start discussing it with somebody. At first you just think they're an imbecile. Then they call you an imbecile, and you want to defend yourself. Crescit eundo. You notice all the little ways they're lying to you and themselves and their audience every time they open their mouth to defend their imbecilic opinion. Then you notice how all the lies are connected, that in order to keep getting the little things like the Scissor statement wrong, they have to drag in everything else. Eventually even that doesn't work, they've just got to make everybody hate you so that nobody will even listen to your argument no matter how obviously true it is. Finally, they don't care about the Scissor statement anymore. They've just dug themselves so deep basing their whole existence around hating you and wanting you to fail that they can't walk it back. You've got to prove them wrong, not because you care about the Scissor statement either, but because otherwise they'll do anything to poison people against you, make it impossible for them to even understand the argument for why you deserve to exist. You know this is true. Your mind becomes a constant loop of arguments you can use to defend yourself, and rehearsals of arguments for why their attacks are cruel and unfair, and the one burning question: how can you thwart them? How can you convince people not to listen to them, before they find those people and exploit their biases and turn them against you? How can you combat the superficial arguments they're deploying, before otherwise good people get convinced, so convinced their mind will be made up and they can never be unconvinced again? *How can you keep yourself safe?* Shiri read two or three sample Scissor statements to me. She didn't say if she agreed with them or not. I didn't tell her if I agreed with them or not. They were harmless.

I don't hear voices in a crazy way. But sometimes I talk to myself. Sometimes I do both halves of the conversation. Sometimes I imagine one of them is a different person. I had a tough breakup a year ago. Sometimes the other voice in my head is my ex-girlfriend's voice. I know how she thinks and I always know what she would say about everything. So sometimes I hold conversations with her, even though she isn't there, and we've barely talked since the breakup. I don't know if this is weird. If it is, I'm weird.

And that was enough. For some reason, it was the third-highest-ranked Scissor statement that did it. None of the others, just that one. The totally hypothetical conversation with the version of my ex-girlfriend in my head about the third Scissor statement got me. Shiri's Scissor was never really about other people anyway. Other people are just the trigger – and I use that word deliberately, in the trigger warning sense. Once you're triggered, you never need to talk

to anyone else again. Just the knowledge that those people are out there is enough.

I thought I'd be done with this story in a night. Instead it's taken me two weeks, all the way up until Halloween – perfect night for a ghost story, right? I've been alternately drinking and smoking weed, trying to calm myself down enough to think about anything other than the third Scissor statement. No, that's not right, definitely trying not to think about either of the first two Scissor statements, because if I think about them, I might start thinking about how some people disagree with them, and then I'm gone. Three times I've started to call my ex-girlfriend to ask her where she is, and if I ever go through with it and she answers me, I don't know what I will do to her. But it isn't just her. Fifty percent of the population disagrees with me on the third-highest-ranked Scissor statement. I don't know who they are. I haven't really appreciated that fact. Not really. I can't imagine it being anyone I know. They're too decent. But I can't be sure it isn't. So I drink.

I know I should be talking about how we all need to unite against whatever shadowy manipulators keep throwing Scissor statements at us. I want to talk about how we need to cultivate radical compassion and charity as the only defense against such abominations. I want to give an Obamaesque speech about how the ties that bring us together are stronger than the forces tearing us apart. But I can't.

Remember what we did to Mozambique? How out of some vestigial sense of ethics, we released a low-potency Scissor statement? Arranged to give them a bad time without destroying the whole country all at once? That's what our shadowy manipulators are doing to us. Low-potency statements. Enough to get us enraged. Not enough to start Armageddon.

But I read the whole list. And then, like an idiot, I thought about it. I thought about the third-highest-ranked Scissor statement in enough detail to let it trigger. To even begin to question whether it might be true is so sick, so perverse, so hateful and disgusting, that Idi Amin would flush with shame to even contemplate it. And if the Scissor's right then half of you would be gung ho in support.

You guys, who haven't heard a really bad Scissor statement yet and don't know what it's like – it's easy for you to say "don't let it manipulate you" or "we need a hard and fast policy of not letting ourselves fight over Scissor statements". But how do you know you're not in the wrong? How do you know there's not an issue out there where, if you knew it, you would agree it would be better to just nuke the world and let us start over again from the sewer mutants, rather than let the sort of people who would support it continue to pollute the world with their presence? How do you know that you're not like the schoolkid who superciliously says "*Nothing* is bad enough to deserve a swear word" when the worst that's ever happened to her is dropping her lollipop in the dirt. If that schoolkid gets kidnapped and tortured, does she change her mind? If she can't describe the torture to her schoolmates, but just says "a really bad thing happened to me", and they still insist nothing could be bad enough to justify using swear words, who do you side with? Then why are you still thinking I'm "damaged" when I tell you I've seen the Scissor statement, and charity and compassion and unity can fuck off and die? Some last remnant of [outside-view morality](#) keeps me from writing the whole list here and letting you all exterminate yourselves. Some remnant of how I would have thought about these things a month ago holds me back. So listen:

Delete Facebook. Delete Twitter. Throw away your cell phone. Unsubscribe from the newspaper. Tell your friends and relatives not to discuss politics or society. If they slip up, break off all contact.

Then, buy canned food. Stockpile water. Learn to shoot a gun. If you can afford a bunker, get a bunker.

Because one day, whoever keeps feeding us Scissor statements is going to release one of the bad ones.

Science and Pseudoscience



“ *The first principle is that you must not fool yourself and you are the easiest person to fool.* ”

— Richard Feynman

☆☆

[\[REPOST\] Epistemic Learned Helplessness](#)

[If It's Worth Doing, It's Worth Doing With Made-Up Statistics](#)

[Lizardman's Constant Is 4%](#)

[The Control Group Is Out Of Control](#)

[Meat Your Doom](#)

[SSC Journal Club: Expert Prediction Of Experiments](#)

[My IRB Nightmare](#)

[Book Review: The Structure Of Scientific Revolutions](#)

[REPOST] Epistemic Learned Helplessness

Posted on June 3, 2019 by Scott Alexander



This is a slightly edited repost of an essay from my old LiveJournal.

**

A friend recently complained about how many people lack 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 envisioned an art of rationality that would make people 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 a terrible idea."

I don't think I'm overselling myself too much to expect that I could argue circles around the average uneducated person. Like I mean that on most topics, I could demolish their position and make them look like an idiot. Reduce them 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. Maybe not on *every* topic, 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 those of the universally reviled crackpots who write books about [Venus being a comet](#).

You could consider this a form of epistemic learned helplessness, where I know any attempt to evaluate the arguments is 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: if I know that a false argument sounds just as convincing as a true argument, argument convincingness provides no evidence either way. 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'll trust the evidence of my own reason. In fact, I trust it in most cases other than infamously deceptive arguments in fields I know little about. But I think the average uneducated person doesn't and shouldn't. Anyone anywhere – politicians, scammy businessmen, smooth-talking romantic partners – would be able to argue them into anything. And so they take the obvious and correct defensive maneuver – they will never let anyone convince them of any belief that sounds "weird".

(and remember that, if you grow up in the right circles, beliefs along the lines of "astrology doesn't work" sound "weird".)

This is starting to resemble ideas like 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, taking ideas seriously is the wrong strategy. Or, at the very least, it 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 resist your natural and detrimental tendency not to take ideas seriously. But I think they might instead co-occur because you have to be really smart in order for taking ideas seriously not to be 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 gets partial credit, at best.

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'd 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 blow myself up."

Responsible doctors are at the other end of the spectrum from terrorists here. 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 – unless I am an expert in that particular subsubfield of medicine, it can sound very convincing even

when it's very wrong.

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. 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.

I'm 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.

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

Posted on May 2, 2013 by Scott Alexander



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."

Lizardman's Constant Is 4%

Posted on April 12, 2013 by Scott Alexander



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 – jesters, 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.

The Control Group Is Out Of Control

Posted on April 28, 2014 by Scott Alexander



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 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 it out.

Luckily we have a natural experiment in terms of parapsychology – the study of psychic phenomena – which most reasonable people believe don't exist, but which a community of practicing scientists believes in 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, $p < 1.2 \times 10^{-10}$, 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. Yet its conclusion is that psychic powers exist.

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 commandments 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, and the researchers here seem to be making some serious mistakes. 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 going 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 sensible 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 Taddio [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 hypothetical 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:	0.268
Ritchie 1:	0.015
Ritchie 2:	'-0.219
Richie 3:	'-0.040
Subbotsky 1:	0.279
Subbotsky 2:	0.292
Subbotsky 3:	'-0.399

Three find large positive effects, two find approximate zero effects, and two find large negative effects. Without doing any calculation, 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 commitments 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 made 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

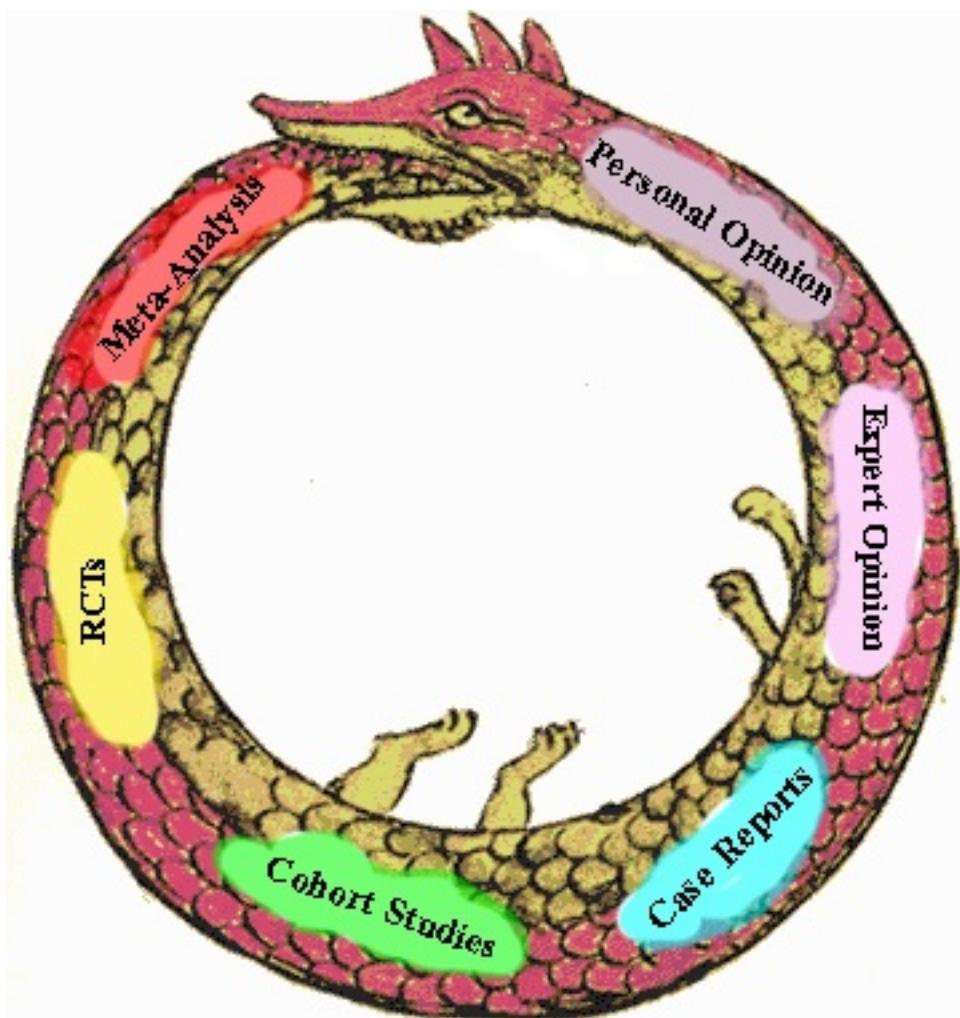
And for a while this went *well*. People would drop things off towers, or see how quickly gases expanded, or observe chimpanzees, or whatever.

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 $p < 1.2 \times 10^{-10}$ 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...



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

Meat Your Doom

Posted on October 28, 2015 by Scott Alexander



Epistemic status: Very dirty and approximate, but I think roughly correct. Check my calculations and tell me if I'm wrong.

I

A recent formative experience: a seriously ill patient came in and I recommended a strong psychiatric drug. She looked it up online and told me she wouldn't take it because was associated with an X% increase in mortality.

"But," I pointed out, "you're really miserable."

"But I don't want to die!"

So I looked it up, did the calculations, and found that it would on average take a couple of months off her life. And I asked her, "Which would you prefer – living 80 years severely ill, or living 79.5 years feeling mostly okay?"

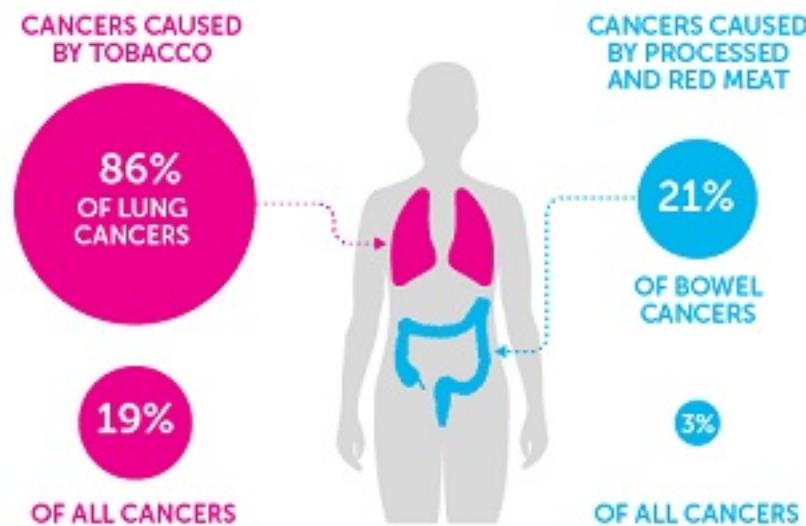
She still wasn't convinced, so I asked her if she ate cookies. She said yes, almost every day. I told her that the cookies were probably taking more time off her life than the medication would, and I assured her the medication would probably add more value to her life than cookies.

She took the drug.

I thought of this the other day when everyone started sharing that study about meat causing colon cancer. A lot of people used headlines like [Processed Meats Rank Alongside Smoking As Cancer Causes](#). This was very correctly debunked by infographics like this one:

TOBACCO vs MEAT WHAT'S THE RISK?

The EVIDENCE that processed meat causes cancer is as strong as the evidence for tobacco, but the RISK from tobacco is much higher...



THE NUMBER OF CANCERS PER YEAR IN THE UK THAT COULD BE PREVENTED IF...



Source: cruk.org/cancerstats



But I feel like this leaves something to be desired. Eating meat is not as bad as smoking. But that's still a lot of room for it to be bad. Can we quantify the risk better?

From [the BBC article](#): “[There would be] one extra case of bowel cancer in 100 lifetime bacon-eaters,’ argues Sir David Spiegelhalter, a risk professor from the University of Cambridge.”

This teaches us something important: “risk professor” is an awesome job title and “David Spiegelhalter, Risk Professor” ought to be a BBC television show starring Harrison Ford.

But also: use absolute risk instead of relative risk! “21% of bowel cancers are caused by meat” doesn’t give you a really good handle on how worried you should be. “One extra case of bowel-cancer in 100 lifetime bacon-eaters” is better.

But let me try to give even more perspective. A bit less than half of colon cancers are fatal. So one extra case per

hundred means if you eat bacon daily then there's an 0.4% chance you will die from a cancer you would not otherwise have gotten.

The average age at diagnosis of colon cancer is 69; the average life expectancy is 79. Sweeping a lot of complexity under the rug and taking a very liberal estimate, the average death from colon cancer costs you ten years of your life.

Multiply out and an 0.4% chance of losing 10 years means that you lose on average two weeks.

Suppose that every case of cancer, fatal and non-fatal alike, causes you additional non-death-related distress equal to two years of your life. That's about another week.

So overall, if you eat processed meat every day your entire life, you'll lose about three weeks of life expectancy from colon cancer. That means each serving of meat costs you a minute of your life. You probably lose twenty times that amount just cooking and preparing it.

II

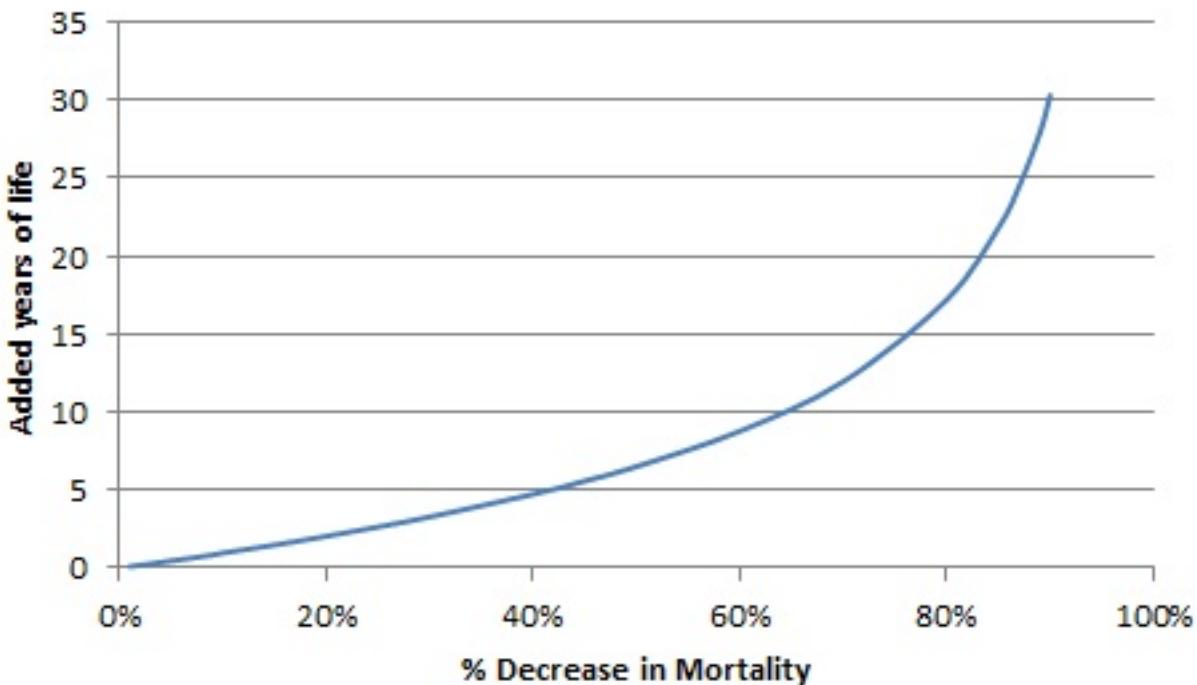
Note that I am not saying "eating meat will only decrease your lifespan by three weeks". That is the amount that we have *clear evidence* for, *from this study*. It is an example of why *this study* needs to be put in context so that you don't worry about it too much.

There are nevertheless a lot of other studies that suggest greater risks, mostly cardiovascular or metabolic. For example, as per [this article](#), some studies suggest that a serving of red meat per day increases mortality 13%, and a serving of processed meat per day increases it 20%. But it also quotes another study of half a million people that finds meat to be slightly *protective* (sigh) and finds a higher all-cause mortality in the non-meat-eaters.

Whatever. Forget the object-level question for a minute. What are we to make of a claim like "processed meat increases mortality 20%"?

If you're like me, you want to think "Okay, average life expectancy is eighty years, subtract 20% off of that, and you get 64 years. I'll live 64 years if I eat bacon every day." WRONG. Mortality rates are much more complicated, but the key insight is that very few people die when they're young. If you have approximately a 0% chance of dying at age 30, then adding 20% to 0 is still 0. Chance of mortality creeps upward *very slowly* and so even large changes in mortality barely affect the underlying distribution. The only good presentation of this I have ever seen anywhere is [on Josh Mitteldorf's blog](#), which includes the following chart:

Cutting Mortality adds Years to Life SLOWLY



This is decrease and we're talking increase, but it shouldn't make much difference here. A 20% increase in mortality isn't going to bring you from 80 to 64. It'll probably just bring you from 80 to 78.

Indeed, later in the BBC article, they bring in David Spiegelhalter (RISK PROFESSOR!) who explains that:

If the studies are right... you would expect someone who eats a bacon sandwich every day to live, on average, two years less than someone who does not. Pro rata, this is like losing an hour of your life for every bacon sandwich you eat. To put this into context, every time you smoke 20 cigarettes, this will take about five hours off your life.

That's for processed meat. Red meat is safer. Also, we still don't know if these studies are right.

This is why it's important to distinguish between absolute and relative risk. You hear all of these scary numbers – 21% increase in bowel cancers! 20% increase in all-cause mortality! – and it sounds like you're going to drop dead the moment you take a bite of a hot dog. And there's always that chance. Being healthy is good. Being unhealthy is bad. But is life so dear or peace so sweet, that you're never going to want to sacrifice an hour to have a bacon sandwich?

All these hours do add up. I'm not saying dietary recommendations aren't important. But the recommendations are important *in aggregate*. If you stick to the spirit of not eating in a horribly unhealthy way, you have a lot of leeway to continue to eat specific things you like even if you know they're not the best for you. And meat falls firmly within that category.

(though you might also want to consider [how to manage the moral issues](#))

SSC Journal Club: Expert Prediction Of Experiments

Posted on November 27, 2016 by Scott Alexander



I

It's been a good month for fretting over failures of expert opinion, so let's look at DellaVigna & Pope, [Predicting Experimental Results: Who Knows What?](#) The authors ran a pretty standard behavioral economics experiment where they asked people on Mechanical Turk to do a boring task while being graded on speed and accuracy. Then they offered one of fifteen different incentive schemes, like "we'll pay you extra if you do well" or "your score will be publicly visible".

But the point of the study wasn't to determine which incentive scheme worked the best, it would determine *who could best predict* which incentive scheme worked the best. The researchers surveyed a bunch of people – economics professors, psychology professors, PhD students, undergrads, business students, and random Internet users on Mechanical Turk – and asked them to predict the experimental results. Since this was a pretty standard sort of behavioral economics experiment, they were wondering whether people with expertise and knowledge in the field might be better than randos at figuring out which schemes would work.

They found that knowledgeable academics had some advantage over randos, but with enough caveats that it's worth going over in more detail.

First, they found that prestigious academics did no better (and possibly slightly worse) than less prestigious academics. Full professors did no better than associate professors, assistant professors, or PhD students. People with many publications and citations did no better than people with fewer publications and citations.

Second, they found that field didn't matter. Behavioral economists did as well as microeconomists did as well as experimental psychologists did as well as theoretical psychologists. To be fair, this experiment was kind of in the intersection of economics and psychology, so all of these fields had equal claim to it. I would have liked to see some geologists or political scientists involved, but they weren't.

Third, the expert advantage was present in one measure of accuracy (absolute forecast error), but not in another (rank-order correlation). On this second measure, experts and randos did about equally well. In other words, experts were better at guessing the exact number for each condition, but not any better at guessing which conditions would do better or worse relative to one another.

Fourth, the expert advantage was pretty small. Professors got an average error of 169, PhD students of 171, undergrads of 187, MBA students of 198, and MTurk users of 271 (random guessing gave an error of about 416). So the difference between undergrads and experts, although statistically significant, was hardly overwhelming.

Fifth, even the slightest use of "wisdom of crowds" was enough to overwhelm the expert advantage. A group of five undergrads averaged together had average error 115, again compared to individual experts' error of 169! Five undergrads averaged together (115) did about as well as five experts averaged together (114). Twenty undergrads averaged together (95) did about as well as twenty experts averaged together (99).

Sixth, having even a little knowledge of individuals' forecasting ability screened off expert status. The researchers gave forecasters some experimental data about the effects of a one-cent incentive and a ten-cent incentive, and asked them to predict the scores after a four-cent incentive – a simple, mechanical problem that just requires common sense. Rандос who can do well on this problem do just as well as experts on the experiment as a whole. Likewise, rандос who are noticed to do well on the first half of the experiment will do just as well as experts on the second half too. In other words, we're back to finding "superforecasters", people who are just consistently good at this kind of thing.

None of this seems to be too confounded by effort. The researchers are able to measure how much time people take on the task, whether they read the instructions carefully, etc. There is some advantage to not rushing through the task, but after that it doesn't seem to matter much. They also try offering some of the Mechanical Turkers lots of money for getting the answers right. That doesn't seem to help much either.

The researchers ask the experts to predict the results of this experiment. They (incorrectly) predict that prestigious academics with full professorships and lots of citations will do better than mere PhD students. They (incorrectly) predict that psychologists will do better than non-psychologists. They (correctly) predict that professors and PhD students will do better than undergrads and rандос.

II

What do we make of this?

I would *tentatively* suggest it doesn't look like experts' expertise is helping them very much here. Part of this is that experts in three different fields did about equally well in predicting the experimental results. But this is only weak evidence; it could be that the necessary expertise is shared among those three fields, or that each field contains one helpful insight and someone who knew all three fields would do better than any of the single-field experts.

But more important, rандос who are able to answer a very simple question, or who do well on other similar problems, do just as well as the experts. This suggests it's possible to get expert-level performance just by being clever, without any particular expertise.

So is it just IQ? This is a tempting explanation. The US average IQ is 100. The undergrads in this experiment came from Berkeley, and Berkeley undergrads have an average SAT of 1375 = average IQ of 133 (this seems really high, but apparently matches estimates from [The Bell Curve](#) and the [Brain Size blog](#); however, see Vaniver's point [here](#)). That same Brain Size post proposes that the average professor has an IQ of 133, but I would expect psychology/economics professors to be higher, plus most of the people in this experiment were from really good schools. If we assume professors are 135-140, then this would neatly predict the differences seen from MTurkers to undergrads to professors.

But the MBA students *really* don't fit into this model. The experiment gets them from the University of Chicago Booth School of Business, which is the top business school in the country and has an average GMAT score of 740. That corresponds to an IQ of almost 150, meaning this should be the highest-IQ sample in the study, yet the MBAs do worse than the undergrads. Unless I'm missing something, this is fatal to an IQ-based explanation.

I think that, as in *Superforecasting*, the best explanation is a separate "rationality" skill which is somewhat predicted by high IQ and scientific training, but not identical to either of them. Although some scientific fields can help you learn the basics of thinking clearly, it doesn't matter what field you're in or whether you're in any field at all as long as you get there somehow.

I'm still confused by the MBA students, and expect to remain so. All MBA students were undergraduates once upon a

time. Most of them probably took at least one economics class, which was where the researchers found and recruited their own undergraduates from. And most of them were probably top students from top institutions, given that they made it into the best business school in the US. So how come Berkeley undergraduates taking an econ class outperform people who *used to be* Berkeley undergraduates taking an econ class, but are now older and wiser and probably a little more selected? It might be that business school selects against the rationality skill, or it might be that business students learn some kind of anti-insight that systematically misleads them in these kinds of problems.

(note that the MBAs don't put in less effort than the other groups; if anything, the reverse pattern is found)

III

Does this relate to interesting real-world issues like people's trouble predicting this election?

One important caveat: this is all atheoretical. As far as I know, there's no theory of psychology or economics that should let people predict how the incentive experiment would go. So it's asking experts to use their intuition, supposedly primed by their expertise, to predict something they have no direct knowledge about. If the experiment were, say, physicists being asked to predict the speed of a falling object, or biologists being asked to predict how quickly a gene with a selective advantage would reach fixation, then we'd be in a very different position.

Another important caveat: predictive tasks are different than interpretative tasks. Ability to predict how an experiment will go without having any data differs from ability to crunch data in a complicated field and conclude that eg saturated fat causes/doesn't cause heart attacks. I worry that a study like this might be used to discredit eg nutritional experts, and to argue that they might not be any better at nutrition than smart laymen. Whether or not this is true, the study doesn't support it.

So one way of looking at it might be that this is a critique not of expertise, but of "punditry". Engineers are still great at building bridges, doctors are still great at curing cancer, physicists are still great at knowing physics – but if you ask someone to predict something vaguely related to their field that they haven't specifically developed and tested a theory to cope with, they won't perform too far above bright undergrads. I think this is an important distinction.

But let's also not get too complacent. The experts in this study clearly *thought* they would do better than PhD students. They *thought* that their professorships and studies and citations would help them. They were wrong. The distinction between punditry and expertise is pretty fuzzy. Had this study come out differently, I could have argued for placing nice clear lab experiments about incentive schemes in the "theory-based and amenable to expertise" category. You can spin a lot of things either direction.

I guess really the only conclusion you can draw from all of this is not to put any important decisions in the hands of people from top business schools.

My IRB Nightmare

Posted on August 29, 2017 by Scott Alexander



Epistemic status: Pieced together from memory years after the event. I may have mis-remembered some things or gotten them in the wrong order. Aside from that – and the obvious jokes – this is all true. I'm being deliberately vague in places because I don't want to condemn anything specific without being able to prove anything.

September 2014

There's a screening test for bipolar disorder. You ask patients a bunch of things like "Do you ever feel really happy, then really sad?". If they say 'yes' to enough of these questions, you start to worry.

Some psychiatrists love this test. I hate it. Patients will say "Yes, that absolutely describes me!" and someone will diagnose them with bipolar disorder. Then if you ask what they meant, they'd say something like "Once my local football team made it to the Super Bowl and I was really happy, but then they lost and I was really sad." I don't even want to tell you how many people get diagnosed bipolar because of stuff like this.

There was a study that supposedly proved this test worked. But parts of it confused me, and it was done on a totally different population that didn't generalize to hospital inpatients. Also, it said in big letters THIS IS JUST A SCREENING TEST IT IS NOT INTENDED FOR DIAGNOSIS, and everyone was using it for diagnosis.

So I complained to some sympathetic doctors and professors, and they asked "Why not do a study?"

Why *not* do a study? Why not join the great tradition of scientists, going back to Galileo and Newton, and make my mark on the world? Why not replace my griping about bipolar screening with an experiment about bipolar screening, an experiment done to the highest standards of the empirical tradition, one that would throw the entire weight of the scientific establishment behind my complaint? I'd been writing about science for so long, even doing my own informal experiments, why not move on to join the big leagues?

For (it would turn out) a whole host of excellent reasons that I was about to learn.

A spring in my step, I journeyed to my hospital's Research Department, hidden in a corner office just outside the orthopaedic ward. It was locked, as always. After enough knocking, a lady finally opened the door and motioned for me to sit down at a paperwork-filled desk.

"I want to do a study," I said.

She looked skeptical. "Have you done the Pre-Study Training?"

I had to admit I hadn't, so off I went. The training was several hours of videos about how the Nazis had done unethical human experiments. Then after World War II, everybody met up and decided to only do ethical human experiments from then on. And the most important part of being ethical was to have all experiments monitored by an Institutional Review Board (IRB) made of important people who could check whether experiments were ethical or not. I dutifully

parroted all this back on the post-test (“Blindly trusting authority to make our ethical decisions for us is the *best* way to separate ourselves from the Nazis!”) and received my Study Investigator Certification.

I went back to the corner office, Study Investigator Certification in hand.

“I want to do a study,” I said.

The lady still looked skeptical. “Do you have a Principal Investigator?”

Mere resident doctors weren’t allowed to do studies on their own. They would probably screw up and start building concentration camps or something. They needed an attending (high-ranking doctor) to sign on as Principal Investigator before the IRB would deign to hear their case.

I knew exactly how to handle this: one by one, I sought out the laziest attendings in the hospital and asked “Hey, would you like to have your name on a study as Principal Investigator for free while I do all the actual work?” Yet one by one, all of the doctors refused, as if I was offering them some kind of plague basket full of vermin. It was the weirdest thing.

Finally, there was only one doctor left – Dr. W, the hardest-working attending I knew, the one who out of some weird masochistic impulse took on every single project anyone asked of him and micromanaged it to perfection, the one who every psychiatrist in the whole hospital (including himself) had diagnosed with obsessive-compulsive personality disorder.

“Sure Scott,” he told me. “I’d be happy to serve as your Principal Investigator”.

A feeling of dread in my stomach, I walked back to the tiny corner office.

“I want to do a study,” I said.

The lady still looked skeptical. “Have you completed the New Study Application?” She gestured to one of the stacks of paperwork filling the room.

It started with a section on my research question. Next was a section on my proposed methodology. A section on possible safety risks. A section on recruitment. A section on consent. A section on... wow. Surely this can’t *all* be the New Study Application? Maybe I accidentally picked up the Found A New Hospital Application?

I asked the lady who worked in the tiny corner office whether, since I was just going to be asking bipolar people whether they ever felt happy and then sad, maybe I could get the short version of the New Study Application?

She told me that *was* the short version.

“But it’s twenty-two pages!”

“You haven’t done any studies before, have you?”

Rather than confess my naivete, I started filling out the twenty-two pages of paperwork. It started by asking about our study design, which was simple: by happy coincidence, I was assigned to Dr. W’s inpatient team for the next three months. When we got patients, I would give them the bipolar screening exam and record the results. Then Dr. W. would conduct a full clinical interview and formally assess them. We’d compare notes and see how often the screening test results matched Dr. W’s expert diagnosis. We usually got about twenty new patients a week; if half of them were willing and able to join our study, we should be able to gather about a hundred data points over the next three months.

It was going to be easy-peasy.

That was the first ten pages or so of the Application. The rest was increasingly bizarre questions such as "Will any organs be removed from participants during this study?" (Look, I promise, I'm not a Nazi.)

And: "Will prisoners be used in the study?" (COME ON, I ALREADY SAID I WASN'T A NAZI.)

And: "What will you do if a participant dies during this research?" (If somebody dies while I'm asking them whether they sometimes feel happy and then sad, I really can't even promise so much as "not freaking out", let alone any sort of dignified research procedure).

And more questions, all along the same lines. I double-dog swore to give everybody really, really good consent forms. I tried my best to write a list of the risks participants were taking upon themselves (mostly getting paper cuts on the consent forms). I argued that these compared favorably to the benefits (maybe doctors will stop giving people strong psychiatric medications just because their football team made the Super Bowl).

When I was done, I went back to the corner office and submitted everything to the Institutional Review Board. Then I sat back and hoped for the best. Like an idiot.

October 2014

The big day arrived. The IRB debated the merits of my study, examined the risks, and... sent me a letter pointing out several irregularities in my consent forms.

IRREGULARITY #1: Consent forms traditionally included the name of the study in big letters where the patient could see it before signing. Mine didn't. Why not?

Well, because in questionnaire-based psychological research, you *never* tell the patient what you're looking for before they fill out the questionnaire. That's like Methods 101. The name of my study was "Validity Of A Screening Instrument For Bipolar Disorder". Tell the patient it's a study about bipolar disorder, and the gig is up.

The IRB listened patiently to my explanation, then told me that this was not a legitimate reason not to put the name of the study in big letters on the consent form. Putting the name of the study on the consent form was important. You know who *else* didn't put the name of the study on his consent forms? *Hitler*.

IRREGULARITY #2: Consent forms traditionally included a paragraph about the possible risks of the study and a justification for why we believed that the benefits were worth the risks. Everyone else included a paragraph about this on our consent forms, and read it to their patients before getting their consent. We didn't have one. Why not?

Well, for one thing, because all we were doing was asking them whether they felt happy and then sad sometimes. This is the sort of thing that goes on every day in a psychiatric hospital. Heck, the other psychiatrists were using this same screening test, except *for real*, and they never had to worry about whether it had risks. In the grand scheme of things, this just wasn't a very risky procedure.

Also, psychiatric patients are sometimes... how can I put this nicely?... a little paranoid. Sometimes you can offer them breakfast and they'll accuse you of trying to poison them. I had no illusions that I would get every single patient to consent to this study, but I felt like I could at least avoid handing them a paper saying "BY THE WAY, THIS STUDY IS FULL OF RISKS".

The IRB listened patiently to my explanation, then told me that this was not a legitimate reason not to have a paragraph about risks. We should figure out some risks, then write a paragraph explaining how those were definitely the risks and we took them very seriously. The other psychiatrists who used this test every day didn't have to do that because they weren't running a study.

IRREGULARITY #3: Signatures are traditionally in pen. But we said our patients would sign in pencil. Why?

Well, because psychiatric patients aren't allowed to have pens in case they stab themselves with them. I don't get why stabbing yourself with a pencil is any less of a problem, but the rules are the rules. We asked the hospital administration for a one-time exemption, to let our patients have pens just long enough to sign the consent form. Hospital administration said absolutely not, and they didn't care if this sabotaged our entire study, it was pencil or nothing.

The IRB listened patiently to all this, then said that it had to be in pen. You know who else had people sign consent forms in pencil...?

I'm definitely not saying that these were the only three issues the IRB sprung on Dr. W and me. I'm saying these are a *representative sample*. I'm saying I spent several weeks relaying increasingly annoyed emails and memos from myself to Dr. W to the IRB to the lady in the corner office to the IRB again. I began to come home later in the evening. My relationships suffered. I started having dreams about being attacked by giant consent forms filled out in pencil.

I was about ready to give up at this point, but Dr. W insisted on combing through various regulations and talking to various people, until he discovered some arcane rule that certain very safe studies with practically no risk were allowed to use an "expedited consent form", which was a lot like a normal consent form but didn't need to have things like the name of the study on it. Faced with someone even more obsessive and bureaucratic than they were, the IRB backed down and gave us preliminary permission to start our study.

The next morning, screening questionnaire in hand, I showed up at the hospital and hoped for the best. Like an idiot.

November 2014

Things progressed slowly. It turns out a lot of psychiatric inpatients are either depressed, agitated, violent, or out of touch with reality, and none of these are really conducive to wanting to participate in studies. A few of them already delusionally thought we were doing experiments on them, and got confused when we suddenly asked them to consent. Several of them made it clear that they hated us and wanted to thwart us in any way possible. After a week, I only had three data points, instead of the ten I'd been banking on.

"Data points" makes it sound abstract. It wasn't. I had hoped to put the results in the patients' easily accessible online chart, *the same place everyone else put the results of the exact same bipolar screening test* when they did it for real. They would put it in a section marked TEST RESULTS, which was there to have a secure place where you could put test results, and where everybody's secure test results were kept.

The IRB would have none of this. Study data are Confidential and need to be kept Secure. Never mind that all the patients' *other* secure test results were on the online chart. Never mind that the online chart contains all sorts of stuff about the patients' diagnoses, medications, hopes and fears, and even (remember, this is a psych hospital) secret fetishes and sexual perversions. Study data needed to be encrypted, then kept in a Study Binder in a locked drawer in a locked room that nobody except the study investigators had access to.

The first problem was that nobody wanted to give us a locked room that nobody except us had access to. There was a

sort of All Purpose Psychiatry Paperwork room, but the janitors went in to clean it out every so often, and apparently this made it unacceptable. Hospitals aren't exactly drowning in spare rooms that not even janitors can get into. Finally Dr. W grudgingly agreed to keep it in his office. This frequently meant I couldn't access any of the study material because Dr. W was having important meetings that couldn't be interrupted by a resident barging into his office to rummage in his locked cabinets.

But whatever. The bigger problem was the encryption. There was a very specific way we had to do it. We would have a Results Log, that said things like "Patient 1 got a score of 11.5 on the test". And then we'd have a Secret Patient Log, which would say things like "Patient 1 = Bob Johnson from Oakburg." That way nobody could steal our results and figure out that Bob was sometimes happy, then sad.

(meanwhile, all of Bob's actual diagnoses, sexual fetishes, etc were in the easily-accessible secure online chart that we were banned from using)

And then – I swear this is true – we had to keep the Results Log and the Secret Patient Log right next to each other in the study binder in the locked drawer in the locked room.

I wasn't sure I was understanding this part right, so I asked Dr. W whether it made sense, to him, that we put a lot of effort writing our results in code, and then put the key to the code in the same place as the enciphered text. He cheerfully agreed this made no sense, but said we had to do it or else our study would fail an audit and get shut down.

January 2015

I'd planned to get a hundred data points in three months. Thanks to constant bureaucratic hurdles, plus patients being less cooperative than I expected, I had about twenty-five. Now I was finishing my rotation on Dr. W's team and going to a clinic far away. What now?

A bunch of newbies were going to be working with Dr. W for the next three months. I hunted them down and threatened and begged them until one of them agreed to keep giving patients the bipolar screening test in exchange for being named as a co-author. Disaster averted, I thought. Like an idiot.

Somehow news of this arrangement reached the lady in the corner office, who asked whether the new investigator had completed her Pre-Study Training. I protested that she wasn't designing the study, she wasn't conducting any analyses, all she was doing was asking her patients the same questions that she would be asking them anyway as part of her job for the next three months. The only difference was that she was recording them and giving them to me.

The lady in the corner office wasn't impressed. You know who *else* hadn't thought his lackeys needed to take courses in research ethics?

So the poor newbie took a course on how Nazis were bad. Now she could help with the study, right?

Wrong. We needed to submit a New Investigator Form to the IRB and wait for their approval.

Two and a half months later, the IRB returned their response: Newbie was good to go. She collected data for the remaining two weeks of her rotation with Dr. W before being sent off to another clinic just like I was.

July 2015

Dr. W and I planned ahead. We had figured out which newbies would be coming in to work for Dr. W three months ahead of time, and gotten them through the don't-be-a-Nazi course and the IRB approval process just in time for them to start their rotation. Success!

Unfortunately, we received another communication from the IRB. Apparently we were allowed to use the expedited consent form to get consent for our *study*, but not to get consent to *access protected health information*. That one required a whole different consent form, list-of-risks and all. We were right back where we'd started from.

I made my case to the Board. My case was: we're not looking at any protected health information, f@#k you.

The Board answered that we were accessing the patient's final diagnosis. It said right in the protocol, we were giving them the screening test, then comparing it to the patient's final diagnosis. "Psychiatric diagnosis" sure *sounds* like protected health information.

I said no, you don't understand, we're the psychiatrists. Dr. W is the one making the final diagnosis. When I'm on Dr. W's team, I'm in the room when he does the diagnostic interview, half the time I'm the one who types the final diagnosis into the chart. These are *our patients*.

The Board said this didn't matter. We, as the patient's doctors, would make the diagnosis and write it down on the chart. But we (as study investigators) needed a full signed consent form before we were allowed to access the diagnosis we had just made.

I said wait, you're telling us we have to do this whole bureaucratic rigamarole with all of these uncooperative patients before we're allowed to see something we wrote ourselves?

The Board said yes, exactly.

I don't remember this part very well, except that I think I half-heartedly trained whichever poor newbie we were using that month in how to take a Protected Health Information Consent on special Protected Health Information Consent Forms, and she nodded her head and said she understood. I think I had kind of clocked out at this point. I was going off to work all the way over in a different town for a year, and I was just sort of desperately hoping that Dr. W and various newbies would take care of things on their own and then in a year when I came back to the hospital I would have a beautiful pile of well-sorted data to analyze. Surely trained doctors would be able to ask simple questions from a screening exam on their own without supervision, I thought. Like an idiot.

July 2016

I returned to my base hospital after a year doing outpatient work in another town. I felt energized, well-rested, and optimistic that the bipolar screening study I had founded so long ago had been prospering in my absence.

Obviously nothing remotely resembling this had happened. Dr. W had vaguely hoped that I was taking care of it. I had vaguely hoped that Dr. W was taking care of it. The various newbies whom we had strategically enlisted had either forgotten about it, half-heartedly screened one or two patients before getting bored, or else mixed up the growing pile of consent forms and releases and logs so thoroughly that we would have to throw out all their work. It had been a year and a half since the study had started, and we had 40 good data points.

The good news was that I was back in town and I could go back to screening patients myself again. Also, we had some particularly enthusiastic newbies who seemed really interested in helping out and getting things right. Over the next three months, our sample size shot up, first to 50, then to 60, finally to 70. Our goal of 100 was almost in sight. The

worst was finally behind me, I hoped. Like an idiot.

November 2016

I got an email saying our study was going to be audited.

It was nothing personal. Some higher-ups in the nationwide hospital system had decided to audit every study in our hospital. We were to gather all our records, submit them to the auditor, and hope for the best.

Dr. W, who was obsessive-compulsive at the best of times, became unbearable. We got into late-night fights over the number of dividers in the study binder. We hunted down every piece of paper that had ever been associated with anyone involved in the study in any way, and almost came to blows over how to organize it. I started working really late. My girlfriend began to doubt I actually existed.

The worst part was all the stuff the newbies had done. Some of them would have the consent sheets numbered in the upper left-hand-corner instead of the upper-right-hand corner. Others would have written the patient name down on the Results Log instead of the Secret Code Log right next to it. One even wrote something in green pen on a formal study document. It was hopeless. Finally we just decided to throw away all their data and pretend it had never existed.

With that decision made, our work actually started to look pretty good. As bad as it was working for an obsessive-compulsive boss in an insane bureaucracy, at least it had the advantage that – when nitpicking push came to ridiculous shove – you were going to be super-ready to be audited. I hoped. Like an idiot.

December 2016

The auditor found twenty-seven infractions.

She was very apologetic about it. She said that was actually a pretty good number of infractions for a study this size, that we were actually doing pretty well compared to a lot of the studies she'd seen. She said she absolutely wasn't going to shut us down, she wasn't even going to censure us. She just wanted us to make twenty-seven changes to our study and get IRB approval for each of them.

I kept the audit report as a souvenir. I have it in front of me now. Here's an example infraction:

The data and safety monitoring plan consists of 'the Principal Investigator will randomly check data integrity'. This is a prospective study with a vulnerable group (mental illness, likely to have diminished capacity, likely to be low income) and, as such, would warrant a more rigorous monitoring plan than what is stated above. In addition to the above, a more adequate plan for this study would also include review of the protocol at regular intervals, on-going checking of any participant complaints or difficulties with the study, monitoring that the approved data variables are the only ones being collected, regular study team meetings to discuss progress and any deviations or unexpected problems. Team meetings help to assure participant protections, adherence to the protocol. Having an adequate monitoring plan is a federal requirement for the approval of a study. See Regulation 45 CFR 46.111 Criteria For IRB Approval Of Research. IRB Policy: PI Qualifications And Responsibility In Conducting Research. Please revise the protocol via a protocol revision request form. Recommend that periodic meetings with the research team occur and be documented.

Among my favorite other infractions:

1. The protocol said we would stop giving the screening exam to patients if they became violent, but failed to rigorously define "violent".
2. We still weren't educating our patients enough about "Alternatives To Participating In This Study". The auditor agreed that the only alternative was "not participating in this study", but said that we had to tell every patient that, then document that we'd done so.
3. The consent forms were still getting signed in pencil. We are never going to live this one down. If I live to be a hundred, representatives from the IRB are going to break into my deathbed room and shout "YOU LET PEOPLE SIGN CONSENT FORMS IN PENCIL, HOW CAN YOU JUSTIFY THAT?!"
4. The woman in the corner office who kept insisting everybody take the Pre-Study Training... hadn't taken the Pre-Study Training, and was therefore unqualified to be our liaison with the IRB. I swear I am not making this up.

Faced with submitting twenty-seven new pieces of paperwork to correct our twenty-seven infractions, Dr. W and I gave up. We shredded the patient data and the Secret Code Log. We told all the newbies they could give up and go home. We submitted the Project Closure Form to the woman in the corner office (who as far as I know still hasn't completed her Pre-Study Training). We told the IRB that they had won, fair and square; we surrendered unconditionally.

They didn't seem the least bit surprised.

August 2017

I've been sitting on this story for a year. I thought it was unwise to publish it while I worked for the hospital in question. I still think it's a great hospital, that it delivers top-notch care, that it has amazing doctors, that it has a really good residency program, and even that the Research Department did everything it could to help me given the legal and regulatory constraints. I don't want this to reflect badly on them in any way. I just thought it was wise to wait a year.

During that year, Dr. W and I worked together on two less ambitious studies, carefully designed not to require any contact with the IRB. One was a case report, the other used publicly available data.

They won 1st and 2nd prize at a regional research competition. I got some nice certificates for my wall and a little prize money. I went on to present one of them at the national meeting of the American Psychiatric Association, a friend helped me write it up formally, and it was recently accepted for publication by a medium-tier journal.

I say this not to boast, but to protest that I'm not as much of a loser as my story probably makes me sound. I'm capable of doing research, I think I have something to contribute to Science. I still think the bipolar screening test is inappropriate for inpatient diagnosis, and I still think that patients are being harmed by people's reliance on it. I still think somebody should look into it and publish the results.

I'm just saying it's not going to be me. I am *done* with research. People keep asking me "You seem really into science, why don't you become a researcher?" Well...

I feel like a study that realistically could have been done by one person in a couple of hours got dragged out into hundreds of hours of paperwork hell for an entire team of miserable doctors. I think its scientific integrity was

screwed up by stupid requirements like the one about breaking blinding, and the patients involved were put through unnecessary trouble by being forced to sign endless consent forms screaming to them about nonexistent risks.

I feel like I was dragged almost to the point of needing to be in a psychiatric hospital myself, while my colleagues who just *used* the bipolar screening test – without making the mistake of trying to check if it works – continue to do so without anybody questioning them or giving them the slightest bit of aggravation.

I feel like some scientists do amazingly crappy studies that couldn't possibly prove anything, but get away with it because they have a well-funded team of clerks and secretaries who handle the paperwork for them. And that I, who was trying to do everything right, got ground down with so many pointless security-theater-style regulations that I'm never going to be able to do the research I would need to show they're wrong.

In the past year or so, I've been gratified to learn some other people are thinking along the same lines. Somebody linked me to [The Censor's Hand](#), a book by a law/medicine professor at the University of Michigan. A summary from [a review](#):

Schneider opens by trying to tally the benefits of IRB review. "Surprisingly," he writes, a careful review of the literature suggests that "research is not especially dangerous. Some biomedical research can be risky, but much of it requires no physical contact with patients and most contact cannot cause serious injury. Ill patients are, if anything, safer in than out of research." As for social-science research, "its risks are trivial compared with daily risks like going online or on a date."

Since the upsides of IRB review are likely to be modest, Schneider argues, it's critical to ask hard questions about the system's costs. And those costs are serious. To a lawyer's eyes, IRBs are strangely unaccountable. They don't have to offer reasons for their decisions, their decisions can't be appealed, and they're barely supervised at the federal level. That lack of accountability, combined with the gauzy ethical principles that govern IRB deliberations, is a recipe for capriciousness. Indeed, in Schneider's estimation, IRBs wield coercive government power—the power to censor university research—without providing due process of law.

And they're not shy about wielding that power. Over time, IRB review has grown more and more intrusive. Not only do IRBs waste thousands of researcher hours on paperwork and elaborate consent forms that most study participants will never understand. Of greater concern, they also superintend research methods to minimize perceived risks. Yet IRB members often aren't experts in the fields they oversee. Indeed, some know little or nothing about research methods at all.

IRBs thus delay, distort, and stifle research, especially research on vulnerable subgroups that may benefit most from it. It's hard to precise about those costs, but they're high: after canvassing the research, Schneider concludes that "IRB regulation annually costs thousands of lives that could have been saved, unmeasurable suffering that could have been softened, and uncountable social ills that could have been ameliorated."

This view seems to be growing more popular lately, and has gotten support from high-profile academics like Richard Nisbett and Steven Pinker:

Should IRBs (human subjects research approval committees) be dismantled? [Probably yes.]
<http://t.co/5mxhEycEA5>

— [Steven Pinker \(@sapinker\)](#), July 24, 2015

And there's been some recent reform, maybe. The federal Office for Human Research Protections [made a vague statement](#) that perhaps studies that obviously aren't going to hurt anybody might not need the full IRB treatment. There's still a lot of debate about how this will be enforced and whether it's going to lead to any real-life changes. But I'm glad people are starting to think more about these things.

(I'm also glad people are starting to agree that getting rid of a little oversight for the lowest-risk studies is a good compromise, and that we don't have to start with anything more radical.)

I sometimes worry that people misunderstand the case against bureaucracy. People imagine it's Big Business complaining about the regulations preventing them from steamrolling over everyone else. That hasn't been my experience. Big Business – heck, Big Anything – loves bureaucracy. They can hire a team of clerks and secretaries and middle managers to fill out all the necessary forms, and the rest of the company can be on their merry way. It's everyone else who suffers. The amateurs, the entrepreneurs, the hobbyists, the people doing something as a labor of love. Wal-Mart is going to keep selling groceries no matter how much paperwork and inspections it takes; the poor immigrant family with the backyard vegetable garden might not.

Bureaucracy in science does the same thing: limit the field to big institutional actors with vested interests. No amount of hassle is going to prevent the Pfizer-Merck-Novartis Corporation from doing whatever study will raise their bottom line. But enough hassle *will* prevent a random psychiatrist at a small community hospital from pursuing his pet theory about bipolar diagnosis. The more hurdles we put up, the more the scientific conversation skews in favor of Pfizer-Merck-Novartis. And the less likely we are to hear little stuff, dissenting voices, and things that don't make anybody any money.

I'm not just talking about IRBs here. I could write a book about this. There are so many privacy and confidentiality restrictions around the most harmless of datasets that research teams won't share data with one another (let alone with unaffiliated citizen scientists) lest they break some arcane regulation or other. Closed access journals require people to pay thousands of dollars in subscription fees before they're allowed to read the scientific literature; open-access journals just shift the burden by requiring scientists to pay thousands of dollars to publish their research. Big research institutions have whole departments to deal with these kinds of problems; unaffiliated people who just want to look into things on their own are out of luck.

And this is happening at the same time we're becoming increasingly aware of the shortcomings of big-name research. [Half of psychology studies](#) fail replication; my own field of psychiatry [is even worse](#). And citizen-scientists and science bloggers are playing a big part in debunking bad research: here I'm thinking especially of statistics bloggers like [Andrew Gelman](#) and [Daniel Lakens](#), but there are all sorts of people in this category. And both Gelman and Lakens are PhDs with institutional affiliations – "citizen science" doesn't mean random cavemen who don't understand the field – but they're both operating outside their day job, trying to contribute a few hours per project instead of a few years. I know many more people like them – smart, highly-qualified, but maybe not going to hire a team of paper-pushers and spend thousands of dollars in fees in order to say what they have to say. Even now these people are doing great work – but I can't help but feel like more is possible.

IRB overreach is a small part of the problem. But it's the part which sunk my bipolar study, a study I really cared about. I'm excited that there's finally more of a national conversation about this kind of thing, and hopeful that further changes will make scientific efforts easier and more rewarding for the next generation of doctors.

Book Review: The Structure Of Scientific Revolutions

Posted on January 8, 2019 by Scott Alexander



I

When I hear scientists talk about Thomas Kuhn, he sounds very reasonable. Scientists have theories that guide their work. Sometimes they run into things their theories can't explain. Then some genius develops a new theory, and scientists are guided by that one. So the cycle repeats, knowledge gained with every step.

When I hear philosophers talk about Thomas Kuhn, he sounds like a madman. There is no such thing as ground-level truth! Only theory! No objective sense-data! Only theory! No basis for accepting or rejecting any theory over any other! Only theory! No scientists! Only theories, wearing lab coats and fake beards, hoping nobody will notice the charade!

I decided to read Kuhn's [*The Structure Of Scientific Revolutions*](#) in order to understand this better. Having finished, I have come to a conclusion: yup, I can see why this book causes so much confusion.

At first Kuhn's thesis appears simple, maybe even obvious. I found myself worrying at times that he was knocking down a straw man, although of course we have to [read the history of philosophy backwards](#) and remember that Kuhn may already be in the water supply, so to speak. He argues against a simplistic view of science in which it is merely the gradual accumulation of facts. So Aristotle discovered a few true facts, Galileo added a few more on, then Newton discovered a few more, and now we have very many facts indeed.

In this model, good science cannot disagree with other good science. You're either wrong – as various pseudoscientists and failed scientists have been throughout history, positing false ideas like "the brain is only there to cool the blood" or "the sun orbits the earth". Or you're right, your ideas are enshrined in the Sacristy Of Settled Science, and your facts join the accumulated store that passes through the ages.

Simple-version-of-Kuhn says this isn't true. Science isn't just facts. It's *paradigms* – whole ways of looking at the world. Without a paradigm, scientists wouldn't know what facts to gather, how to collect them, or what to do with them once they had them. With a paradigm, scientists gather and process facts in the ways the paradigm suggests ("normal science"). Eventually, this process runs into a hitch – apparent contradictions, or things that don't quite fit predictions, or just a giant ugly mess of epicycles. Some genius develops a new paradigm ("paradigm shift" or "scientific revolution"). Then the process begins again. Facts can be accumulated within a paradigm. And many of the facts accumulated in one paradigm can survive, with only slight translation effort, into a new paradigm. But scientific progress is the story of one relatively-successful and genuinely-scientific effort giving way to a different and contradictory relatively-successful and genuinely-scientific effort. It's the story of scientists constantly tossing out one another's work and beginning anew.

This gets awkward because paradigms look a lot like facts. The atomic theory – the current paradigm in a lot of chemistry – looks a lot like the fact "everything is made of atoms and molecules". But this is only the iceberg's tip. Once you have atomic theory, chemistry starts looking a lot different. Your first question when confronted with an unknown chemical is "what is the molecular structure?" and you have pretty good ideas for how to figure this out. You are not particularly interested in the surface appearance of chemicals, since you know that iron and silver can look alike but

are totally different elements; you may be much more interested in the weight ratio at which two chemicals react (which might seem to the uninitiated like a pretty random and silly thing to care about). If confronted with a gas, you might ask things like “which gas is it?” as opposed to thinking all gases are the same thing, or wondering what it would even mean for two gases to be different. You can even think things like “this is a mixture of two different types of gas” without agonizing about how a perfectly uniform substance can be a mixture of anything. If someone asks you “How noble and close to God would say this chemical sample is?” you can tell them that this is not really a legitimate chemical question, unless you mean “noble” in the sense of the noble gases. If someone tells you a certain chemical is toxic because toxicity is a fundamental property of its essence, you can tell them that no, it probably has to do with some reaction it causes or fails to cause with chemicals in the body. And if someone tells you that a certain chemical has changed into a different chemical because it got colder, you can tell them that cold might have done something to it, it might even have caused it to react with the air or something, but chemicals don’t change into other chemicals in a fundamental way just because of the temperature. None of these things are obvious. All of them are hard-won discoveries.

A field without paradigms looks like the STEM supremacist’s stereotype of philosophy. There are all kinds of different schools – Kantians, Aristotelians, Lockeans – who all disagree with each other. There may be progress within a school – some Aristotelian may come up with a really cool new Aristotelian way to look at bioethics, and all the other Aristotelians may agree that it’s great – but the field as a whole does not progress. People will talk past one another; the Aristotelian can go on all day about the telos of the embryo, but the utilitarian is just going to ask what the hell a telos is, why anyone would think embryos have one, and how many utils the embryo is bringing people. “Debates” between the Aristotelian and the utilitarian may not be literally impossible, but they are going to have to go all the way to first principles, in a way that never works. Kuhn interestingly dismisses these areas as “the fields where people write books” – if you want to say anything, you might as well address it to a popular audience for all the good other people’s pre-existing knowledge will do you, and you may have to spend hundreds of pages explaining your entire system from the ground up. He throws all the social sciences in this bin – you may read Freud, Skinner, and Beck instead of Aristotle, Locke, and Kant, but it’s the same situation.

A real science is one where everyone agrees on a single paradigm. Newtonianism and Einsteinianism are the same kind of things as Aristotelianism and utilitarianism; but in 1850, everybody believed the former, and in 1950, the latter.

I got confused by this – is Aristotelian philosophy a science? Would it be one if the Aristotelians forced every non-Aristotelian philosopher out of the academy, so that 100% of philosophers fell in line behind Aristotle? I *think* Kuhn’s answer to this is that it’s telling that Aristotelians haven’t been able to do this (at least not lately); either Aristotle’s theories are too weak, or philosophy too intractable. But all physicists unite behind Einstein in a way that all philosophers cannot behind Aristotle. Because of this, all physicists mean more or less the same thing when they talk about “space” and “time”, and they can work together on explaining these concepts without constantly arguing to each other about what they mean or whether they’re the right way to think about things at all (and a Newtonian and Einsteinian would *not* be able to do this with each other, any more than an Aristotelian and utilitarian).

So how does science settle on a single paradigm when other fields can’t? Is this the part where we admit it’s because science has objective truth so you can just settle questions with experiments?

This is very much not that part. Kuhn doesn’t think it’s anywhere near that simple, for a few reasons.

First, there is rarely a single experiment that one paradigm fails and another passes. Rather, there are dozens of experiments. One paradigm does better on some, the other paradigm does better on others, and everyone argues over which ones should or shouldn’t count.

For example, one might try to test the Copernican vs. Ptolemaic worldviews by observing the parallax of the fixed stars

over the course of a year. Copernicus predicts it should be visible; Ptolemy predicts it shouldn't be. It isn't, which means either the Earth is fixed and unmoving, or the stars are unutterably unimaginably immensely impossibly far away. Nobody expected the stars to be that far away, so advantage Ptolemy. Meanwhile, the Copernicans posit far-off stars in order to save their paradigm. What looked like a test to select one paradigm or the other has turned into a wedge pushing the two paradigms even further apart.

What looks like a decisive victory to one side may look like random noise to another. Did you know weird technologically advanced artifacts [are sometimes found](#) encased in rocks that our current understanding of geology says are millions of years old? Creationists have no trouble explaining those – the rocks are much younger, and the artifacts were probably planted by nephilim. Evolutionists have no idea how to explain those, and default to things like “the artifacts are hoaxes” or “the miners were really careless and a screw slipped from their pocket into the rock vein while they were mining”. I’m an evolutionist and I agree the artifacts are probably hoaxes or mistakes, even when there is no particular evidence that they are. Meanwhile, probably creationists say that some fossil or other incompatible with creationism is a hoax or a mistake. But that means the “find something predicted by one paradigm but not the other, and then the failed theory comes crashing down” oversimplification doesn’t work. Find something predicted by one paradigm but not the other, and often the proponents of the disadvantaged paradigm can – and should – just shrug and say “whatever”.

In 1870, flat-earther Samuel Rowbotham performed [a series of experiments](#) to show the Earth could not be a globe. In the most famous, he placed several flags miles apart along a perfectly straight canal. Then he looked through a telescope and was able to see all of them in a row, even though the furthest should have been hidden by the Earth’s curvature. Having done so, he concluded the Earth was flat, and the spherical-earth paradigm debunked. Alfred Wallace (more famous for pre-empting Darwin on evolution) took up the challenge, and showed that the bending of light rays by atmospheric refraction explained Rowbotham’s result. It turns out that light rays curve downward at a rate equal to the curvature of the Earth’s surface! Luckily for Wallace, refraction was already a known phenomenon; if not, it would have been the same kind of wedge-between-paradigms as the Copernicans having to change the distance to the fixed stars.

It is all nice and well to say “Sure, it *looks* like your paradigm is right, but once we adjust for this new idea about the distance to the stars / the refraction of light, the evidence actually supports my paradigm”. But the supporters of old paradigms can do that too! The Ptolemaics are rightly mocked for adding epicycle after epicycle until their system gave the right result. But to a hostile observer, positing refraction effects that exactly counterbalance the curvature of the Earth sure looks like adding epicycles. At some point a new paradigm will win out, and its “epicycles” will look like perfectly reasonable adjustments for [reality’s surprising amount of detail](#). And the old paradigm will lose, and its “epicycles” will look like obvious kludges to cover up that it never really worked. Before that happens... well, good luck.

Second, two paradigms may not even address or care about the same questions.

Let’s go back to utilitarianism vs. Aristotelianism. Many people associate utilitarianism with the [trolley problem](#), which is indeed a good way to think about some of the issues involved. It might be tempting for a utilitarian to think of Aristotelian ethics as having some different answer to the trolley problem. Maybe it does, I don’t know. But Aristotle doesn’t talk about how he would solve whatever the 4th-century BC equivalent of the trolley problem was. He talks more about “what is the true meaning of justice?” and stuff like that. While you can twist Aristotle into having an opinion on trolleys, he’s not really optimizing for that. And while you can make utilitarianism have some idea what the true meaning of justice is, it’s not really optimized for that either.

An Aristotelian can say their paradigm is best, because it does a great job explicating all the little types and subtypes of justice. A utilitarian can say *their* paradigm is best, because it does a great job telling you how to act in various contrived moral dilemmas.

It's actually even worse than this. The closest thing I can think of to an ancient Greek moral dilemma is the story of Antigone. Antigone's uncle declares that her traitorous dead brother may not be buried with the proper rites. Antigone is torn between her duty to obey her uncle, and her desire to honor her dead brother. Utilitarianism is... not really designed for this sort of moral dilemma. Is ignoring her family squabbles and trying to cure typhus an option? No?

But then utilitarianism's problems are deeper than just "comes to a different conclusion than ancient Greek morals would have". The utilitarian's job isn't to change the ancient Greek's mind about the answer to a certain problem. It's to convince him to stop caring about basically all the problems he cares about, and care about different problems instead.

Third, two paradigms may disagree on what kind of answers are allowed, or what counts as solving a problem.

Kuhn talks about the 17th century "dormitive potency" discourse. Aristotle tended to explain phenomena by appealing to essences; trees grew because it was "in their nature" to grow. Descartes gets a bad rap for inventing dualism, but this is undeserved – what he was really doing was inventing the concept of "matter" as we understand it, a what-you-see-is-what-you-get kind of stuff with no hidden essences, which responds mechanically to forces (and once you have this idea, you naturally need some other kind of stuff to be the mind). With Cartesian matter firmly in place, everyone made fun of Aristotle for thinking he had "solved" the "why do trees grow?" question by answering "because it is in their nature", and this climaxed with the playwright Molire portraying a buffoonish doctor who claimed to have discovered how opium put people to sleep – it was because it had a dormitive potency!

In Aristotle's view of matter, saying "because it's their essence" successfully answers questions like "why do trees grow?". The Cartesian paradigm forbade this kind of answer, and so many previously "solved" problems like why trees grow became mysterious again – a step backwards, sort of. For Descartes, you were only allowed to answer questions if you could explain how purely-mechanical matter smashing against other purely-mechanical matter in a billiard-ball-like way could produce an effect; a more virtuous and Descartes-aware doctor explained opium's properties by saying opium corpuscles must have a sandpaper-like shape that smooths the neurons!

Then Newton discovered gravity and caused an uproar. Gravity posits no corpuscles jostling other corpuscles. It sounds almost Aristotelian: "It is the nature of matter to attract other matter". Newton was denounced as trying to smuggle occultism into science. How much do you discount a theory for having occult elements? If some conception of quantum theory predicts the data beautifully, but says matter behaves differently depending on whether someone's watching it or not, is that okay? What if it says that a certain electron has a 50% chance of being in a certain place, full stop, and there is no conceivable explanation for which of the two possibilities is realized, and you're not even allowed to ask the question? What if my explanation for dark matter is "invisible gremlins"? How do you figure out when you need to relax your assumptions about what counts as science, versus when somebody is just cheating?

A less dramatic example: Lavoisier's theory of combustion boasts an ability to explain why some substances gain weight when burned; they are absorbing oxygen from the air. A brilliant example of an anomaly explained, which proves the superiority of combustion theory to other paradigms that cannot account for the phenomenon? No – "things shouldn't randomly gain weight" comes to us as a principle of the chemical revolution of which Lavoisier was a part:

In the seventeenth century, [an explanation of weight gain] seemed unnecessary to most chemists. If chemical reactions could alter the volume, color, and texture of the ingredients, why should they not alter weight as well? Weight was not always taken to be the measure of quantity of matter. Besides, weight-gain on roasting remained an isolated phenomenon. Most natural bodies (eg wood) lose weight on roasting as the phlogiston theory was later to say they should.

In previous paradigms, weight gain wasn't even an anomaly to be explained. It was just a perfectly okay thing that

might happen. It's only within the constellation of new methods and rules we learned around Lavoisier's time, that Lavoisier's theories solved anything at all.

So how *do* scientists ever switch paradigms?

Kuhn thinks it's kind of an ugly process. It starts with exasperation; the old paradigm is clearly inadequate. Progress is stagnating.

Awareness [of the inadequacy of geocentric astronomy] did come. By the thirteenth century Alfonso X could proclaim that if God had consulted him when creating the universe, he would have received good advice. In the sixteenth century, Copernicus' coworker, Domenico da Novara, held that no system so cumbersome and inaccurate as the Ptolemaic had become could possibly be true of nature. And Copernicus himself wrote in the Preface to the *De Revolutionibus* that the astronomical tradition he inherited had finally created only a monster.

Then someone proposes a new paradigm. In its original form, it is woefully underspecified, bad at matching reality, and only beats the old paradigm in a few test cases. For whatever reason, a few people jump on board. Sometimes the new paradigm is simply more mathematically elegant, more beautiful. Other times it's petty things, like a Frenchman invented the old paradigm and a German the new one, and you're German. Sometimes it's just that there's nothing better. These people gradually expand the new paradigm to cover more and more cases. At some point, the new paradigm explains things a little better than the old paradigm. Some of its predictions are spookily good. The old paradigm is never conclusively debunked. But the new paradigm now has enough advantages that more and more people hop on the bandwagon. Gradually the old paradigm becomes a laughingstock, people forget the context in which it ever made sense, and it is remembered only as a bunch of jokes about dormitive potency.

But now that it's been adopted and expanded and reached the zenith of its power, *this* is the point at which we can admit it's objectively better, right?

For a better treatment of this question than I can give, see Samzdat's [Science Cannot Count To Red](#). But my impression is that Kuhn is not really willing to say this. I think he is of the "all models are wrong, some are useful" camp, thinks of paradigms as models, and would be willing to admit a new paradigm may be more useful than an old one.

Can we separate the fact around which a paradigm is based (like "the Earth orbits the sun") from the paradigm itself (being a collection of definitions of eg "planet" and "orbit", ways of thinking, mathematical methods, and rules for what kind of science will and won't be accepted)? And then say the earth factually orbits the sun, and the paradigm is just a useful tool that shouldn't be judged objectively? I think Kuhn's answer is that facts cannot be paradigm-independent. A medieval would not hear "the Earth orbits the sun" and hear the same claim we hear (albeit, in his view wrong). He would, for example, interpret it to mean the Earth was set in a slowly-turning crystal sphere with the sun at its center. Then he might ask – where does the sphere intersect the Earth? How come we can't see it? Is Marco Polo going to try to travel to China and then hit a huge invisible wall halfway across the Himalayas? And what about gravity? My understanding is the Ptolemaics didn't believe in gravity as we understand it at all. They believed objects had a natural tendency to seek the center of the universe. So if the sun is more central, why isn't everything falling into the sun? To a medieval the statement "the Earth orbits the sun" has a bunch of common-sense disproofs everywhere you look. It's only when attached to the rest of the Copernican paradigm that it starts to make sense.

This impresses me less than it impresses Kuhn. I would say "if you have many false beliefs, then true statements may be confusing in that they seem to imply false statements – but true statements are still objectively true". Perhaps I am misunderstanding Kuhn's argument here; the above is an amalgam of various things and not something Kuhn says outright in the book. But whatever his argument, Kuhn is not really willing to say that there are definite paradigm-independent objective facts, at least not without a lot of caveats.

So where *is* the point at which we admit some things are objectively true and that's what this whole enterprise rests on?

Kuhn only barely touches on this, in the last page of the book:

Anyone who has followed the argument this far will nevertheless feel the need to ask why the evolutionary process should work. What must nature, including man, be like in order that science be possible at all? Why should scientific communities be able to reach a firm consensus unattainable in other fields? Why should consensus endure across one paradigm change after another? And why should paradigm change invariably produce an instrument more perfect in any sense than those known before? From one point of view those questions, excepting the first, have already been answered. But from another they are as open as they were when this essay began. It is not only the scientific community that must be special. The world of which that community is a part must also possess quite special characteristics, and we are no closer than we were at the start to knowing what these must be. That problem—What must the world be like in order that man may know it?—was not, however, created by this essay. On the contrary, it is as old as science itself, and it remains unanswered. But it need not be answered in this place.

II

A lot of the examples above are mine, not Kuhn's. Some of them even come from philosophy or other nonscientific fields. Shouldn't I have used the book's own examples?

Yes. But one of my big complaints about this book is that, for a purported description of How Science Everywhere Is Always Practiced, it really just gives five examples. Ptolemy/Copernicus on astronomy. Alchemy/Dalton on chemistry. Phlogiston/Lavoisier on combustion. Aristotle/Galileo/Newton/Einstein on motion. And ???/Franklin/Coulomb on electricity.

It doesn't explain any of the examples. If you don't already know what Coulomb's contribution to electricity is and what previous ideas he overturned, you're out of luck. And don't try looking it up in a book either. Kuhn says that all the books have been written by people so engrossed in the current paradigm that they unconsciously jam past scientists into it, removing all evidence of paradigm shift. This made parts of the book a little beyond my level, since my knowledge of Coulomb begins and ends with "one amp times one second".

Even saying Kuhn has five examples is giving him too much credit. He usually brings in one of his five per point he's trying to make, meaning that you never get a really full view of how any of the five examples exactly fit into his system.

And all five examples are from physics. Kuhn says at the beginning that he wished he had time to talk about how his system fits biology, but he doesn't. He's unsure whether any of the social sciences are sciences at all, and nothing else even gets mentioned. This means we have to figure out how Kuhn's theory fits everything from scattershot looks at the history of electricity and astronomy and a few other things. This is pretty hard. For example, consider three scientific papers I've looked at on this blog recently:

- [Cipriani, Ioannidis, et al](#) perform a meta-analysis of antidepressant effect sizes and find that although almost all of them seem to work, amitriptyline works best.
- [Ceballos, Ehrlich, et al](#) calculate whether more species have become extinct recently than would be expected based on historical background rates; after finding almost 500 extinctions since 1900, they conclude they definitely have.

- [Terrell et al](#) examine contributions to open source projects and find that men are more likely to be accepted than women when adjusted for some measure of competence they believe is appropriate, suggesting a gender bias.

What paradigm is each of these working from?

You could argue that the antidepressant study is working off of the “biological psychiatry” paradigm, a venerable collection of assumptions that can be profitably contrasted with other paradigms like psychoanalysis. But couldn’t a Hippocratic four-humors physician of a thousand years ago done the same thing? A meta-analysis of the effect sizes of various kinds of leeches for depression? Sure, leeches are different from antidepressants, but it doesn’t look like the belief in biological psychiatry is affecting anything about the research other than the topic. And although the topic is certainly important, Kuhn led me to expect something more profound than that. Maybe the paradigm is evidence-based-medicine itself, the practice of doing RCTs and meta-analyses on things? I think this is a stronger case, but a paradigm completely divorced from the content of what it’s studying is exactly the sort of weird thing that makes me wish Kuhn had included more than five examples.

As for the extinction paper, surely it can be attributed to some chain of thought starting with Cuvier’s catastrophism, passing through Lyell, and continuing on to the current day, based on the idea that the world has changed dramatically over its history and new species can arise and old ones disappear. But is that “the” paradigm of biology, or ecology, or whatever field Ceballos and Lyell are working in? Doesn’t it also depend on the idea of species, a different paradigm starting with Linnaeus and developed by zoologists over the ensuing centuries? It looks like it dips into a bunch of different paradigms, but is not wholly within any.

And the open source paper? Is “feminism” a paradigm? But surely this is no different than what would be done to investigate racist biases in open source. Or some right-winger looking for anti-Christian biases in open source. Is the paradigm just “looking for biases in things?”

What about my favorite trivial example, [looking both ways when you cross the street so you don't get hit by a bus?](#) Is it based on a paradigm of motorized transportation? Does it use assumptions like “buses exist” and “roads are there to be crossed”? Was there a paradigm shift between the bad old days of looking one way before crossing, and the exciting new development of looking both ways before crossing? Is this really that much more of a stretch than calling looking for biases in things a paradigm?

Outside the five examples Kuhn gives from the physical sciences, identifying paradigms seems pretty hard – or maybe too easy. Is it all fractal? Are there overarching paradigms like atomic theory, and then lower-level paradigms like organic chemistry, and then tiny subsubparadigms like “how we deal with this one organic compound”? Does every scientific experiment use lots of different paradigms from different traditions and different levels? This is the kind of thing I wish Kuhn’s book answered instead of just talking about Coulomb and Copernicus over and over again.

III

In conclusion, all of this is about predictive coding.

It’s the same thing. Perception getting guided equally by top-down expectations and bottom-up evidence. Oh, I know what you’re thinking. “There goes Scott again, seeing predictive coding in everything”. And yes. But also, Kuhn does everything short of come out and say “When you guys get around to inventing predictive coding, make sure to notice that’s what I was getting at this whole time.”

Don’t believe me? From the chapter *Anomaly And The Emergence Of Scientific Discovery* (my emphasis, and for

"anomaly", read "surprisal"):

The characteristics common to the three examples above are characteristic of all discoveries from which new sorts of phenomena emerge. Those characteristics include: the previous awareness of anomaly, the gradual and simultaneous emergence of both observational and conceptual recognition, and the consequent change of paradigm categories and procedures often accompanied by resistance. **There is even evidence that these same characteristics are built into the nature of the perceptual process itself.** In a psychological experiment that deserves to be far better known outside the trade, Bruner and Postman asked experimental subjects to identify on short and controlled exposure a series of playing cards. Many of the cards were normal, but some were made anomalous, e.g., a red six of spades and a black four of hearts. Each experimental run was constituted by the display of a single card to a single subject in a series of gradually increased exposures. After each exposure the subject was asked what he had seen, and the run was terminated by two successive correct identifications.

Even on the shortest exposures many subjects identified most of the cards, and after a small increase all the subjects identified them all. For the normal cards these identifications were usually correct, but the anomalous cards were almost always identified, without apparent hesitation or puzzlement, as normal. The black four of hearts might, for example, be identified as the four of either spades or hearts. Without any awareness of trouble, it was immediately fitted to one of the conceptual categories prepared by prior experience. One would not even like to say that the subjects had seen something different from what they identified. With a further increase of exposure to the anomalous cards, subjects did begin to hesitate and to display awareness of anomaly. Exposed, for example, to the red six of spades, some would say: That's the six of spades, but there's something wrong with it—the black has a red border. Further increase of exposure resulted in still more hesitation and confusion until finally, and sometimes quite suddenly, most subjects would produce the correct identification without hesitation. Moreover, after doing this with two or three of the anomalous cards, they would have little further difficulty with the others. A few subjects, however, were never able to make the requisite adjustment of their categories. Even at forty times the average exposure required to recognize normal cards for what they were, more than 10 per cent of the anomalous cards were not correctly identified. And the subjects who then failed often experienced acute personal distress. One of them exclaimed: "I can't make the suit out, whatever it is. It didn't even look like a card that time. I don't know what color it is now or whether it's a spade or a heart. I'm not even sure now what a spade looks like. My God!" In the next section we shall occasionally see scientists behaving this way too.

Either as a metaphor or **because it reflects the nature of the mind**, that psychological experiment provides a wonderfully simple and cogent schema for the process of scientific discovery.

And from *Revolutions As Changes Of World-View*:

Surveying the rich experimental literature from which these examples are drawn makes one suspect that something like a paradigm is prerequisite to perception itself. What a man sees depends both upon what he looks at and also upon what his previous visual-conceptual experience has taught him to see. In the absence of such training there can only be, in William James's phrase, "a bloomin' buzzin' confusion." In recent years several of those concerned with the history of science have found the sorts of experiments described above immensely suggestive.

If you can read those paragraphs and honestly still think I'm just irrationally reading predictive coding into a perfectly innocent book, I have nothing to say to you.

I think this is my best answer to the whole “is Kuhn denying an objective reality” issue. If Kuhn and the predictive coding people are grasping at the same thing from different angles, then both shed some light on each other. I think I understand the way that predictive coding balances the importance of pre-existing structures and categories with a preserved belief in objectivity. If Kuhn is trying to extend the predictive coding model of the brain processing information to the way the scientific community as a whole processes it, then maybe we can import the same balance and not worry about it as much.

Psychiatry, Psychology, & Neurology



“ Korzybski had accepted that ‘...all we know is a joint phenomenon of the observer and the observed...’ One could no longer study how people know what they presume they know (epistemology) without scientific understanding (at a given date) of the ‘physical’, ‘neurobiological’, ‘psycho-social’, ‘linguistic’, etc., factors affecting human observers.”

— Bruce Kodish, *Korzybski: A Biography*

**

The Phatic And The Anti-Inductive
CBT In The Water Supply
[What Universal Human Experiences Are You Missing Without Realizing It?](#)
[What Developmental Milestones Are You Missing?](#)
An Iron Curtain Has Descended Upon Psychopharmacology
[Why Were Early Psychedelicists So Weird?](#)
[Universal Love, Said The Cactus Person](#)
[Did A Melatonin Patent Inspire Current Dose Confusion?](#)
[It's Bayes All The Way Up](#)
[Book Review: Surfing Uncertainty](#)
[Book Review: Behavior – The Control Of Perception](#)
[Predictive Processing And Perceptual Control](#)

The Phatic And The Anti-Inductive

Posted on January 11, 2015 by Scott Alexander



I

Ozy recently taught me the word “phatic”. It means talking for the sake of talking.

The classic example is small talk. “Hey.” “Hey.” “How are you?” “Fine, and you?” “Fine.” No information has been exchanged. Even if the person involved wasn’t fine, they’d still say *fine*. Indeed, at least in this country giving an information-bearing response to “how are you?” is a mild social faux pas.

Some people call this “social grooming behavior” and it makes sense. It’s just a way of saying “Hello, I acknowledge you and still consider you an acquaintance. There’s nothing wrong between us. Carry on.” That you are willing to spend ten seconds holding a useless conversation with them signals this just fine.

We can go a little more complex. Imagine I’m calling a friend from college after five years out of contact; I’ve heard he’s got a company now and I want to ask him for a job. It starts off “Hey, how are you?”, segues into “And how are the wife and kids?”, then maybe into “What are you doing with yourself these days?” and finally “Hey, I have a big favor to ask you.” If you pick up the phone and say “Hello, it’s Scott from college, can you help me get a job?” this is rude. It probably sounds like you’re *using* him.

And I mean, you are. If I cared about him deeply as a person I probably would have called him at some point in the last five years, before I needed something. But by mutual consent we both sweep that under the rug by having a few minutes of meaningless personal conversation beforehand. The information exchanged doesn’t matter – “how’s your business going?” is just as good as “how’s your wife and kids?” is just as good as “how are your parents doing?”. The point is to clock a certain number of minutes about something vaguely personal, so that the request seems less abrupt.

We can go even more complex. By the broadest definition, phatic communication is equivalent to signaling.

Consider a very formulaic conservative radio show. Every week, the host talks about some scandal that liberals have been involved in. Then she explains why it means the country is going to hell. I don’t think the listeners really care that a school in Vermont has banned Christmas decorations or whatever. The point is to convey this vague undercurrent of “Hey, there are other people out there who think like you, we all agree with you, you’re a good person, you can just sit here and listen and feel reassured that you’re right.” Anything vaguely conservative in content will be equally effective, regardless of whether the listener cares about the particular issue.

II

Douglas Adams once said there was a theory that if anyone ever understood the Universe, it would disappear and be replaced by something even more incomprehensible. He added that there was another theory that this had already happened.

These sorts of things – things such that if you understand them, they get more complicated until you don’t – are called “anti-inductive”.

The classic anti-inductive institution is the stock market. Suppose you found a pattern in the stock market. For example, it always went down on Tuesdays, then up on Wednesdays. Then you could buy lots of stock Tuesday evening, when it was low, and sell it Wednesday, when it was high, and be assured of making free money.

But lots of people want free money, so lots of people will try this plan. There will be so much demand for stock on Tuesday evening that there won’t be enough stocks to fill it all. Desperate buyers will bid up the prices. Meanwhile, on Wednesday, everyone will sell their stocks at once, causing a huge glut and making prices go down. This will continue until the trend of low prices Tuesday, high prices Wednesday disappears.

So in general, it should be impossible to exploit your pattern-finding ability to profit of the stock market unless you are the smartest and most resourceful person in the world. That is, maybe stocks go up every time the Fed cuts interest rates, but Goldman Sachs knows that too, so they probably have computers programmed to buy so much stock milliseconds after the interest rate announcement is made that the prices will stabilize on that alone. That means that unless you can predict better than, or respond faster than, Goldman Sachs, you can’t exploit your knowledge of this pattern and shouldn’t even try.

Here’s something I haven’t heard described as anti-inductive before: job-seeking.

When I was applying for medical residencies, I asked some people in the field to help me out with my interviewing skills.

“Why did you want to become a doctor?” they asked.

“I want to help people,” I said.

“Oh God,” they answered. “No, anything but that. Nothing says ‘person exactly like every other bright-eyed naive new doctor’ than wanting to help people. You’re trying to distinguish yourself from the pack!”

“Then... uh... I want to hurt people?”

“Okay, tell you what. You have any experience treating people in disaster-prone Third World countries?”

“I worked at a hospital in Haiti after the earthquake there.”

“Perfect. That’s inspirational as hell. Talk about how you want to become a doctor because the people of Haiti taught you so much.”

Wanting to help people is a great reason to become a doctor. When Hippocrates was taking his first students, he was probably really impressed by the one guy who said he wanted to help people. But since that time it’s become cliche, overused. Now it signals people who can’t come up with an original answer. So you need something better.

During my interviews, I talked about my time working in Haiti. I got to talk to some of the other applicants, and they talked about *their* time working in Ethiopia, or Bangladesh, or Nicaragua, or wherever. Apparently the “stand out by working in a disaster-prone Third World country” plan was sufficiently successful that everyone started using, and now the people who do it don’t stand out at all. My interviewer was probably thinking “Oh God, what Third World country is *this* guy going to start blabbering about how much he learned from?” and moving my application to the REJECT pile as soon as I opened my mouth.

I am getting the same vibe from the critiques of OkCupid profiles in the last open thread. OkCupid seems very susceptible to everybody posting identical quirky pictures of themselves rock-climbing, then talking about how fun-loving and down-to-earth they are. On the other hand, every deviation from that medium has *also* been explored.

"I'm going for 'quirky yet kind'."

"Done."

"Sarcastic, yet nerdy?"

"Done."

"Outdoorsy, yet intellectual."

"Done."

"Introverted, yet a zombie."

"I thought we went over this. [Zombies. Are. Super. Done.](#)"

III

I've been thinking about this lately in the context of psychotherapy.

I'm not talking about the very specific therapies, the ones where they teach special cognitive skills, or expose you to spiders to cure your arachnophobia. They don't let me do those yet. I'm talking about what's called "supportive therapy", where you're just talking to people and trying to make them feel generally better.

When I was first starting out, I tried to do therapy anti-inductively. I figured that I had to come up with something unexpected, something that the patient hadn't thought of. Some kind of brilliant interpretation that put all of their problems in a new light. This went poorly. It tended to be a lot of "Well, have you tried [obvious thing?]", them saying they had, and me escalating to "Well, have you tried [long shot that probably wouldn't work]?"

(I wonder if this was Freud's strategy: "Okay, he says he's depressed, I can't just tell him to cheer up, probably everybody says that. Can't just tell him to accept his sadness, that one's obvious too. Got to come up with something really original... uh..." HAVE YOU CONSIDERED THAT YOU WANT TO KILL YOUR FATHER AND MARRY YOUR MOTHER??!"

Now I tend more to phatic therapy. This happened kind of by accident. Some manic people have a symptom called "pressured speech" which means they never shut up and they never let you get a word in edgewise. Eventually, more out of surrender than out of a strategic plan, I gave up and stopped trying. I just let them talk, nodded my head, said "Yeah, that sounds bad" when they said something bad-sounding, said "Oh, that's good" when they said something good-sounding.

After a while I realized this went at least as well as any other therapy I was doing, plus the patients really liked me and thought I was great and gave me lots of compliments.

So after that, "active listening" became sort of my default position for supportive therapy. Get people talking. Let them

talk. Nod my head as if I am deeply concerned about their problems. Accept their effusive praise about how well I seem to be understanding them.

This is clearly phatic. I would say the ritual is “High status person is willing to listen to my problems. That means society considers my problems important and considers me important. It means my problems are okay to have and I’m not in trouble for having them.” As long as I seem vaguely approving, the ritual reaches its predetermined conclusion.

IV

I was thinking about this recently several friends have told me how much she hated “therapist speak”. You know, things like “I feel your pain” or “And how does that make you feel?”

I interpret this as an anti-inductive perspective on therapy. The first therapist to say “I feel your pain” may have impressed her patients – a person who herself can actually feel all my hurt and anger! Amazing! But this became such a standard in the profession that it became the Default Therapist Response. Now it’s a signal of “I care so little about your pain that I can’t even bother to say anything other than the default response.” When a therapist says “I feel your pain,” it’s easy to imagine that in her head she’s actually planning what she’s going to make for dinner or something.

So just as some people find it useful to divide the world into “ask culture” and “guess culture”, I am finding it useful to divide the world into “phatic culture” and “anti-inductive culture”.

There are people for whom “I feel your pain” is exactly the right response. It shows that you are sticking to your therapist script, it urges them to stick to their patient script, and at the end of the session they feel like the ritual has been completed and they feel better.

There are other people for whom “I feel your pain” is the most enraging thing you could possibly say. It shows that you’re not taking them seriously or engaging with them, just saying exactly the same thing you do to all your other patients.

There are people for whom coming up with some sort of unique perspective or clever solution for their problems is exactly the right response. Even if it doesn’t work, it at least proves that you are thinking hard about what they are saying.

There are other people for whom coming up with some sort of unique perspective or clever solution is the most enraging thing you could possibly do. At the risk of perpetuating gender stereotypes, one of the most frequently repeated pieces of relationship advice I hear is “When a woman is telling you her problems, just listen and sympathize, don’t try to propose solutions”. It sounds like the hypothetical woman in this advice is looking for a phatic answer.

I think myself and most of my friends fall far to the anti-inductive side, with little tolerance for the phatic side. And I think we probably typical-mind other people as doing the same.

This seems related to the classic geek discomfort with small-talk, with pep rallies, and with normal object-level politics. I think it might also be part of the problem I had with social skills when I was younger – I remember talking to people, panicking because I couldn’t think of any way to make the conversation unusually entertaining or enlightening, and feeling like I had been a failure for responding to the boring-weather-related question with a boring-weather-related answer. Very speculatively, I think it might have something to do with creepy romantic overtures – imagine the same mental pattern that made me jokingly consider giving “I want to hurt people” as my motivation for becoming a doctor, applied to a domain that I really don’t understand on a fundamental enough level to know whether

or not saying that is a good idea.

I've been trying to learn the skill of appreciating the phatic. I used to be very bad at sending out thank-you cards, because I figured if I sent a thank-you card that just said "Thank you for the gift, I really appreciate it" then they would think that the lack of personalization meant I wasn't *really* thankful. But personalizing a bunch of messages to people I often don't really know that well is hard and I ended up all miserable. Now I just send out the thank you card with the impersonal message, and most people are like "Oh, it was so nice of you to send me a card, I can tell you really appreciated it." This seems like an improvement.

As for psychotherapy, I think I'm going to default to phatic in most cases when I don't have some incredibly enlightening insight, then let my patients tell me if that's the wrong thing to do.

CBT In The Water Supply

Posted on July 16, 2015 by Scott Alexander



Epistemic status: Very speculative, <50% confidence, thinking out loud. Don't let this turn you off therapy.

**

Here's a vignette from cognitive-behavioral therapy book [When Panic Attacks](#), heavily edited for length:

A chronically anxious medical school professor named Nate suffered from low-self-esteem and feelings of inadequacy. One day, Nate brought me a copy of his CV. I was blown away. He'd listed over sixty pages of research publications, prestigious awards, and keynote addresses he'd given at major conferences around the world. I asked Nate how he reconciled his low self-esteem with all of his accomplishments. He said that every time he looked at his CV, he felt discouraged and told himself that his colleagues' research studies were far more rigorous and important than his own. He said his paper seemed "soft" and consisted primarily of theoretical work, rather than hard-core laboratory research with real tissue. He said "Dr. Burns, no matter how much I accomplish, it never seems good enough."

Perfectionism was clearly one of Nate's self-defeating beliefs. I suggested that Nate use the Pleasure/Perfection Balance Worksheet to test this belief. I told him to write "If I can't do something perfectly, it's not worth doing at all" on the top of the sheet, and asked him to list several activities in the left-hand column. I told him to predict how satisfying and rewarding each activity would be, to record how satisfying and rewarding it was afterwards, and to rate how perfectly he did each activity. That way he could find out if it was true that he only enjoyed the things he did perfectly.

The next week, Nate had some interesting results to share with me. One of his activities was giving the welcoming lecture of the incoming class of medical students. Nate gave this lecture every year because he was considered to be the most charismatic speaker at the medical school. Nate predicted this lecture would be 70% satisfying, but his actual satisfaction was only 20%. This was surprising, since he'd received a thirty-second standing ovation, and he'd rated his perfection level for the talk at 90%.

I asked Nate why his satisfaction rating was so low. He explained that he always got standing ovations, so he routinely timed them. The previous year, the medical students had stood and cheered for more than a minute at the end of his talk. This year, they only stood and cheered for half a minute. Nate felt disappointed and started worrying that he was over the hill.

The second entry on Nate's Pleasure/Perfection Balance Worksheet was that [he fixed a broken pipe in his bathroom]. He had to make several trips to the hardware store to buy tools and parts and to get tips on how to do it, so he didn't get the pipe fixed until 10 PM. He explained that any plumber could have fixed the pipe in five minutes, so he rated his perfection as 5%. But his satisfaction level for this activity was 100%. In fact, he felt exhilarated. Nate said it was the most satisfying thing he'd done in years.

The result of Nate's experiment was not consistent with his belief that things weren't worth doing unless he did them perfectly. It dawned on him that there were many sources of satisfaction in his life that he'd overlooked, such as taking a walk through the woods with his wife, even though neither of them were

world-class hikers, playing squash with his son, even though neither of them were champions, or just going out with his family for ice cream cones on a warm summer evening.

This experiment had a significant impact on Nate's feelings of self-esteem and on his career. He told me that his feelings of anxiety and inferiority decreased, and his productivity actually increased because he was no longer so worried about having to do everything so perfectly.

At first I assumed this story was made up, but the book claims these are based on real patients, and even mentions how the writer showed videos of some of these therapy sessions to his classes. Interesting. How about another?

Several years ago, I did a three-day intensive workshop for a small group of psychotherapists in Florida. A marriage and family therapist named Walter explained that he'd been struggling with anxiety and depression for several months because Paul, the man he'd lived with for eight years, had found a new lover and left him. He put his hand on his chest and said: "It feels real heavy, right here. There's just a sense of loneliness and emptiness about the whole experience. It feels so universal and final. I feel like this pain is going to go on forever, until the end of time."

I asked Walter how he was thinking and feeling about the breakup with Paul. What was he telling himself? He said "I feel incredibly guilty and ashamed, and it seems like it must have been my fault. Maybe I wasn't skillful enough, attractive enough, or dynamic enough. Maybe I wasn't there for him emotionally. I feel like I must have screwed up. Sometimes I feel like a total fraud. Here I am, a marriage and family therapist, and my own relationship didn't even work out. I feel like a loser. A really, really big loser."

Walter recorded these five negative thoughts on his daily mood log:

1. I'll never be in a loving relationship again
2. I must be impossible to live with and impossible to be in a relationship with
3. There must be something wrong with me
4. I totally screwed up and flushed my life down the toilet
5. I'll end up as an old, fat, gray-haired, lonely gay man

He believed all of these thoughts very strongly.

You can see that most of Walter's suffering results from the illogical way he's thinking about the rejection. You could even say that Walter is treating himself far more harshly than Paul did. I thought the Double Standard Technique might help because Walter seemed to be a warm and compassionate individual. I asked what he'd say to a dear friend who'd been rejected by someone he'd been living with for eight years. I said "Would you tell him that there's something wrong with him, that he screwed up his life and flushed it down the toilet for good?"

Walter looked shocked and said he'd *never* say something like that to a friend. I suggested we try a role-playing exercise so that he could tell me what he would say to a friend who was in the same predicament [...]

Therapist (role-playing patient's friend): Walter, there's another angle I haven't told you about. What you don't understand is that I'm impossible to live with and be in a relationship with. That's the real reason I feel so bad, and that's why I'll be alone for the rest of my life.

Patient (role-playing as if therapist is his friend who just had a bad breakup): Gosh, I'm surprised to hear you say that, because I've known you for a long time and never felt that way about you. In fact, you've always been warm and open, and a loyal friend. How in the world did you come to the conclusion that you were impossible to be in a relationship with?

Therapist (continuing role-play): Well, my relationship with [my boyfriend] fell apart. Doesn't that prove I'm impossible to be in a relationship with?

Patient (continuing role-play): In all honesty, what you're saying doesn't make a lot of sense. In the first place, your boyfriend was also involved in the relationship. It takes two to tango. And in the second place, you were involved in a reasonably successful relationship with him for eight years. So how can you claim that you're impossible to live with?

Therapist (continuing role-play:) Let me make sure I've got this right. You're saying that I was in a reasonably successful relationship for eight years, so it doesn't make much sense to say that I'm impossible to live with or impossible to be in a relationship with?

Patient (continuing-role-play:) You've got it. Crystal clear.

At that point, Walter's face lit up, as if a lightbulb had suddenly turned on in his brain, and we both started laughing. His negative thoughts suddenly seemed absurd to him, and there was an immediate shift in his mood... after Walter put the lie to his negative thoughts, I asked him to rate how he was feeling again. His feeling of sadness fell all the way from 80% to 20%. His feelings of guilt, shame, and anxiety fell all the way to 10%, and his feelings of hopelessness dropped to 5%. The feelings of loneliness, embarrassment, frustration, and anger disappeared completely.

The book is quite long, and it's full of stories like this. The author, who's one of the top cognitive-behavioral psychiatrists in the world, describes his experience with the therapy as:

[When I first learned about this therapy, I thought] depression and anxiety seemed far too serious and severe for such a simplistic approach. But when I tried these methods with some of my more difficult patients, my perceptions changed. Patients who'd felt hopeless, worthless, and desperate began to recover. At first, it was hard to believe that the techniques were working, but I could not deny the fact that when my patients learned to put the lie to their negative thoughts, they began to improve. Sometimes they recovered right before my eyes during sessions. Patients who'd felt demoralized and hopeless for years suddenly turned the corner on their problems. I can still recall an elderly French woman who'd been bitterly depressed for more than fifty years, with three nearly-successful suicide attempts, who started shouting "Joie de vivre! Joie de vivre!" ("joy of living") one day in my office. These experiences made such a strong impact on me that I decided my calling was in clinical work rather than brain research. After considerable soul-searching, I decided to give up my research career and become a full-time clinician. Over the years, I've had more than 35,000 psychotherapy sessions with depressed and anxious patients, and I'm every bit as enthusiastic about CBT as when I first began learning about it.

Okay. I am not one of the top cognitive-behavioral therapists in the world. I've been studying formal cognitive-behavioral therapy for about a week now, and been doing untrained ad hoc therapy on inpatients for a couple years. But I've also gotten to observe a lot of other people doing therapy, and talked to people who have had therapy, and treated patients who were simultaneously undergoing therapy, and the impression I got was very different.

Dr. Burns asks patients to question whether their anxiety and their negative thoughts are rational, and their faces light

up and all of their psychiatric problems suddenly melt away.

The therapists I've seen ask patients to question whether their anxiety and their negative thoughts are rational, ever so tactfully, and the patients say "No shit, Sherlock, of course they aren't, but just knowing that doesn't help or make them go away, and I've been through this same spiel with like thirty people already. Now shut up and give me my Xanax."

In my last post, [someone asked](#) what to do if they found cognitive-behavioral therapy hokey and patronizing. I said, only half joking, that "if you don't like hokey patronizing things, CBT may not be for you." I know it's mean, and pessimistic, but everyone I've talked to has had pretty much the same experience. I used to attribute this to my friends being pretty smart, and maybe CBT was aimed at less intelligent people, but Nate The Genius Medical School Professor seems pretty smart. So does Walter The Therapist. Burns' book includes a bunch of other vignettes about high-powered lawyers, graduate students, et cetera. They all find his suggestions of "Well, have you considered that your irrational negative thoughts might not be rational?" super life-changing.

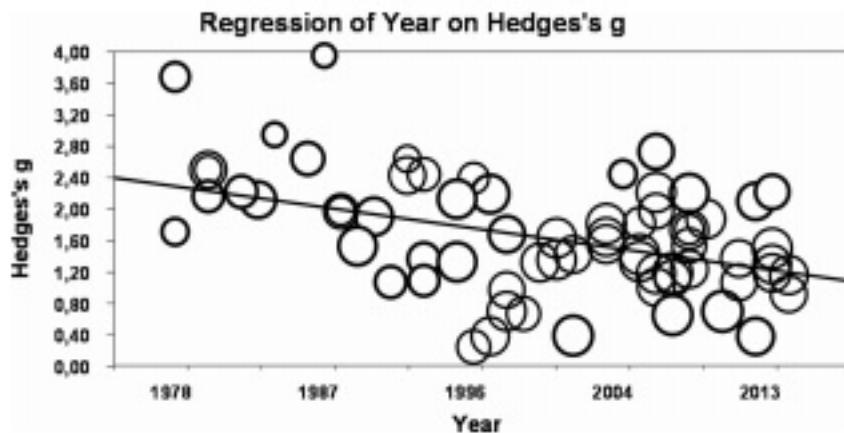


Figure 4. The plot portrays the negative change ($p < .001$) in Beck Depression Inventory effect sizes across time ($k = 61$). The size of the circles indicates the relative contribution (random weight) of each study to the analysis.

You might have read the study this graph comes from: [The Effects of Cognitive-Behavioral Therapy As An Anti-Depressive Treatment Is Falling: A Meta-Analysis](#). As you can see, the [Hedges' g](#) declined from about 2.5 in 1980 to around 1 today. The latest embarrassing set of results now [show](#) CBT doing no better than its old nemesis psychoanalysis. Why?

There are a lot of possible explanations. The smart money is always on "it never worked very well, but we're finally doing studies that aren't hopelessly biased", but the analysis doesn't find a clear difference in study quality. Other suggestions are that therapists have gotten less committed over time, or that the patient populations have changed. All of these sound reasonable. But let me mention one more possibility.

Every so often, psychiatrists joke about how so many people are depressed we might as well put Prozac in the water supply. Sometimes we say the same thing about lithium, although in that case [we're not joking](#).

Nobody's ever talked about putting cognitive-behavioral therapy in the water supply, but insofar as that's meaningful at all I would say we've kind of done it. Cognitive-behavioral ideas, like perfectionism, excessive self-blame,

conditional versus unconditional self-respect, deep breathing, goal-setting, et cetera have become basic parts of popular culture. The whole self-esteem movement isn't *exactly* cognitive-behavioral, but it's certainly allied, and it certainly represents a shift to a style of thinking about the self and about psychology in a way that's much more fertile for cognitive-behavioral ideas. Inside Out was *kind of* "Cognitive Behavioral Therapy: The Movie".

Although the particular book I'm reading is from 2006, Burns himself was one of Aaron Beck's original students and one of the first cognitive-behavioral therapists ever. I wonder how many of these patients who seem absolutely *shocked* to realize that maybe their anxiety isn't rational come from that very early period.

It's very hard to track changes in people's basic beliefs about psychology. I was flabbergasted to learn that until Dr. Benjamin Spock's landmark 1940s book on child care, parents were told not to hug, kiss, or show affection to babies, because that would coddle them and make them weak, pampered adults. Before that, parents interacted with their kids much less, and it was assumed that siblings and nannies and friends would raise them, or they would raise themselves. It's easy to read books about ancient Greece and not notice that they have a completely different view of the role of the self/individual than we do. So it wouldn't surprise me if a lot of the psychology we consider "obvious" is CBT that has seeped out into the water supply over the past thirty years.

If that were true, it would explain why CBT is no longer as effective – it's just telling people things they already know.

It could be fairly asked: then why isn't everybody *already* better? Depression seems to be increasing, though there's a lot of argument about exactly how much; that doesn't sound like what would happen if everyone were automatically getting a background level of therapy.

Here's a theory, though it's on even shakier ground than the other one. The meta-analysis proposes that CBT may have lost some placebo effect over time because patients no longer think of it as The Exciting New Thing. I'm not sure I can go along with that – [my own analysis](#) of psychotropic medications suggests patients very much prefer the *old* ones for some reason. But a big part of psychotherapy [is placebo effect](#), so they might be on to something.

What part of psychotherapy provides the placebo? Is it going to the clinic? Talking to the therapist? Hearing fancy words like "self-estimation"? Doing worksheets?

One thing a lot of therapies have in common is that they provide the feeling of insights. For example, psychoanalysts are very good at coming up with surprising-but-plausible ways that your current problems are linked to things that happened to you as a child; the usual result is a patient feeling enlightened, like "You're right, the leg pain that's been bothering me *is* in the same part of my leg that accidentally brushed up against my mother's breast one time when I was seven, that's pretty interesting."

Suppose that in the old days, CBT was an insight a minute and you were constantly hearing surprising things you'd never thought about before. And nowadays, you're kind of absorbing a lot of those things by osmosis without it seeming too insightful, and then the therapy itself is anticlimactic. Could that lessen the placebo effect enough to account for the data?

I don't know. Maybe after I've been training in formal CBT for more than a week, I'll have more data and can report back to you.

EDIT: Sarah [writes](#): "*In a way, seeing CBT stuff in pop culture inoculates people, I think. People will get as far as noticing "this negative thought is an anxiety symptom", but not as far as *actually reversing it*. When people hadn't heard of CBT, they first got the "this negative thought is irrational" message in a context when they were actively working on their problems, so they followed through with the 'hard' step of actually reversing the thought. Now, people run into the revelation that the 'inner critic' is wrong just by browsing facebook, when they're *not* actively trying to fight their*

anxiety problems, so the revelation loses its force."

EDIT 2: Paul Crowley points out [a very similar theory](#) in *The Guardian*

What Universal Human Experiences Are You Missing Without Realizing It?

Posted on March 17, 2014 by Scott Alexander



Remember [Galton's experiments on visual imagination](#)? Some people just don't have it. And they never figured it out. They assumed no one had it, and when people talked about being able to picture objects in their minds, they were speaking metaphorically.

And the people who *did* have good visual imaginations didn't catch them. The people without imaginations mastered this "metaphorical way of talking" so well that they passed for normal. No one figured it out until Galton sat everyone down together and said "Hey, can we be really really clear about exactly how literal we're being here?" and everyone realized they were describing different experiences.

I thought about this recently during a conversation with Ozy:

Ozy: I am currently eating chickpeas and rice and I am *delighted* by the fact that I can eat this *whenever I want*. The nice thing about DISCOVERING YOUR FOOD PREFERENCES is that suddenly all the food in my cupboards is food I like and am looking forward to eating. And usually I get food I like by, like, luck? So this is excitement.

Scott: I don't understand, why didn't you buy things like that before?

Ozy: It took me a while to have enough of a sense of the food I like for "make a list of the food I like" to be a viable grocery-list-making strategy.

Scott: I've got to admit I'm confused and intrigued by your "don't know my own preferences" thing.

Ozy: Hrm. Well, it's sort of like... you know how sometimes you pretend to like something because it's high-status, and if you do it well enough you *actually believe* you like the thing? Unless I pay a lot of attention *all* my preferences end up being not "what I actually enjoy" but like "what is high status" or "what will keep people from getting angry at me"

Scott: How does that apply to food?

Ozy: Well, sometimes people will tell you a certain food is high-status or healthy or a thing that everyone enjoys, and then I would like it. And a lot of times I just ate whatever was in front of me or ordered whatever the cheapest vegetarian thing on the menu was. And I... sort of vaguely had a sense that some things were more pleasurable to eat than other things but I didn't like *keep track* of what they were or anything. Because if I knew I might like the *wrong things*. And also because I didn't intuitively grasp that the "liking" thing everyone was talking about was related to pleasure and not to like popularity/status.

So the fact that people talk about what foods they like about a zillion times a day isn't enough to make everyone realize liking foods is a thing.

But it gets worse. A high school friend posted on Facebook a link to [a really interesting answer on Quora](#). It makes you log on, so I'll copy the relevant part below:

I have anosmia, which means I lack smell the way a blind person lacks sight. What's surprising about this is that I didn't even know it for the first half of my life.

Each night I would tell my mom, "Dinner smells great!" I teased my sister about her stinky feet. I held my nose when I ate Brussels sprouts. In gardens, I bent down and took a whiff of the roses. I yelled "gross" when someone farted. I never thought twice about any of it for fourteen years.

Then, in freshman English class, I had an assignment to write about the Garden of Eden using details from all five senses. Working on this one night, I sat in my room imagining a peach. I watched the juice ooze out as I squeezed at the soft fuzz. I felt the wet, sappy liquid drip from my fingers down onto my palm. As the mushy heart of the fruit compressed, I could hear it squishing, and when I took that first bite I could taste the little bit of tartness that followed the incredible sweet sensation flooding my mouth.

But I had to write about smell, too, and I was stopped dead by the question of what a peach smelled like. Good. That was all I could come up with. I tried to think of other things. Garbage smelled bad. Perfume smelled good. Popcorn good. Poop bad. But how so? What was the difference? What were the nuances? In just a few minutes' reflection I realized that, despite years of believing the contrary, I never had and never would smell a peach.

All my behavior to that point indicated that I had smell. No one suspected I didn't. For years I simply hadn't known what it was that was supposed to be there. I just thought the way it was for me was how it was for everyone. It took the right stimulus before I finally discovered the gap.

So I guess you can just *not be able to smell* and not know it.

This makes me wonder what universal human experiences I and my friends are missing out on without realizing it.

I know one friend's answer. He discovered he was color-blind sometime in his teens. This still surprises me. People are always taking Ishihara tests (those colorful dotted circles with numbers inside of them) and discovering they're color blind. Going through life with everyone else saying "The light was red, but now it's green" and thinking it was weird that they were making such a big deal about subtle variations in shades of brownish-gray, but it was probably one of those metaphors.

As for me? I took a surprisingly long time to realize I was asexual. When I was a virgin, I figured sex was one of those things that seemed gross before you did it, and then you realized how great it was. Afterwards, I figured it was something that didn't get good until you were skilled at it and had been in a relationship long enough to truly appreciate the other person. In retrospect, pretty much every aspect of male sexual culture is a counterargument to that theory, but I guess it's just really hard for my brain to generate "you are a mental mutant" as a hypothesis.

But even bigger than that, I think I might not have had emotions, at least not fully, for about five years as a teenager when I was on SSRIs. I even sort of noticed myself not having emotions, but dismissed that as an odd thing to happen and probably other people were just being really overexuberant about things. Later I learned emotional blunting is a commonly reported side effect of SSRIs and I was probably just really not experiencing emotions. When I came off them it took me several years to get used to having normal-intensity feelings again, but it wasn't a sudden revelation, like "Wow, I was missing a fundamental human experience for the past several years!" Just a sense of things being different which was hard to cash out.

As always, I wonder if a lot of what other people interpret through vague social things might be biological, or at least more *complicatedly* social. I can't enjoy jazz music even a little – the best I can do is pick up something sort of like a beat and half-heartedly feel like maybe I could snap my fingers to it if I could build up the energy. My brother fell in love with jazz as soon as he heard it and is now a professional jazz musician who has dedicated his life to it. Are we listening to the same thing when we hear a jazz tune? Or am I like a guy who can't smell trying to appreciate perfume?

What Developmental Milestones Are You Missing?

Posted on November 3, 2015 by Scott Alexander



Epistemic status: Speculative. I can't make this post less condescending and elitist, so if you don't like condescending elitist things, this might not be for you.

*
*

Developmental psychology never struck my interest in the same way as a lot of other kinds of psychology. It didn't seem to give me insight into my own life, help me understand my friends, or explain weird things about society.

I've changed my mind about all of that after reading David Chapman's [Developing Ethical, Social, and Cognitive Competence](#).

First, a refresher. Developmental psychology describes how children go from helpless infants to reasonable adults. Although a lot of it has to do with sensorimotor skills like walking and talking, the really interesting stuff is cognitive development. Children start off as very buggy reasoners incapable of all but the most superficial forms of logic but gradually go on to develop new abilities and insights that allow them to navigate adult life.

Maybe the most famous of these is "theory of mind", the ability to view things from other people's perspective. In a classic demonstration, researchers show little Amy a Skittles bag and ask what she thinks is inside. She guesses Skittles, but the researchers open it and reveal it's actually pennies. Then they close it up and invite little Brayden into the room. Then they ask Amy what Brayden thinks is inside. If Amy's three years old or younger, she'll usually say "pennies" – she knows that pennies are inside, so why shouldn't Brayden know too? If she's four or older, she'll usually say "Skittles" – she realizes on a gut level that she and Brayden are separate minds and that Brayden will have his own perspective. Sometimes the same mistake can extend to preferences and beliefs. Wikipedia gives the example of a child saying "I like Sesame Street, so Daddy must like Sesame Street too." This is another theory of mind failure grounded in an inability to separate self and environment.

Here's another example which tentatively sounds like a self-environment failure. Young children *really* don't get foreign languages. I got a little of this teaching English in Japan, and heard more of it from other people. The really young kids treated English like a cipher; everybody started out knowing things' *real* (ie Japanese) names, but Americans insisted on converting them into their own special American-person code before talking about them. Kids would ask weird things like whether American parents would make an exception and speak Japanese to their kids who were too young to have learned English yet, or whether it was a zero-tolerance policy sort of thing and the families would just not communicate until the kids went to English school. And I made fun of them, but I also remember the first time I visited Paris I heard somebody talking to their dog, and for a split second I was like "Why would you expect your dog to know French?" before my brain kicked in and I was like "Duuhhhh..."

The infamous "magical thinking" which kids display until age 7 or so also involves confused self-environment boundaries. Maybe little Amy gets mad at Brayden and shouts "I HATE HIM" to her mother. The next day, Brayden falls off a step and skins his knee. Amy intuits a cause-and-effect relationship between her hatred and Brayden's accident and feels guilty. She doesn't realize that her hatred is internal to herself and can't affect the world directly. Or kids displaying animism at this age, and expecting that the TV doesn't work because it's angry, or the car's not starting because it's tired.

Psychology textbooks never discuss whether this progression in and out of developmental stages is innate or environmental, which is weird because psychology textbooks usually love that sort of thing. I always assumed it was innate, because it was on the same timeline as things like walking and talking which are definitely innate. But I've been moved to question that after reading some of the work comparing "primitive" cultures to primitive developmental stages.

This probably isn't the most politically correct thing to do, but it's notable enough that anthropologists have been thinking about it for centuries. For example, from [*Ethnicity, Nationality, and Religious Experience*](#):

Primitive people are generally as intelligent as the people of any culture, including the contemporary industrial-electronic age cultures. That makes it all the more significant that their publicly shared cognitive style shows little identifiable formal operational thought. The probable explanation for this, if true, is simply that formal operational thought is more complexly difficult than earlier modes of thought and will be used in a culture in a publicly shared way only if that culture has developed techniques for training people in its use. Primitive cultures do not do that, and thus by default use easier styles of thought, ones closer in form to concrete operational and even pre-operational thought, as defined by Piaget.

Primitive cultures certainly exhibit the magical thinking typical of young children; this is the origin of a whole host of superstitions and witch-doctory. They exhibit the same animism; there are hundreds of different animistic religions worldwide. And although I didn't talk much about theories of moral development, primitive cultures' notion of taboo is pretty similar to Kohlberg's conventional stage.

But if different cultures progress through developmental milestones at different rates or not at all, then these aren't universal laws of child development but facts about what skills get learned slowly or quickly in different cultures. In this model, development is not a matter of certain innate abilities like walking "unfolding" at the right time, but about difficult mental operations that you either learn or you don't depending on how hard the world is trying to cram them into your head.

So getting back to David Chapman: his post is mostly about Robert Kegan's account of "stages of moral development". I didn't get much from Kegan himself, but I was fascinated by an idea just sort of dropped into the middle of the discussion: that less than half of the people in modern western countries had attained Kegan's fourth stage, and only a small handful attained his fifth. This was a way of thinking about development that I'd never heard before.

On the other hand, it makes sense. Take General Semantics (please!). I remember reading through Korzybski's [giant blue book of General Semantics](#), full of labyrinthine diagrams and promises that if only you understood this, you would engage with the world totally differently, you'd be a new man armed with invincible cognitive weapons. And the key insight, maybe the *only* insight, was "the map is not the territory", which seems utterly banal.

But this is a self-environment distinction of exactly the sort that children learn in development. It's dividing your own representation of the world from the world itself; it's about as clear a reference to theory of mind as you could ask for. Korzybski considered it a revelation when he discovered it; thousands of other people found it helpful and started a movement around it; I conclude that these people were missing a piece of theory-of-mind and Korzybski gave it to them. Not the whole deal, of course. Just a piece. But a piece of something big and fundamental, so abstract and difficult to teach that it required that whole nine-hundred-something page book to cram it in.

And now I'm looking for other things in the discourse that sound like developmental milestones, and *there are oodles of them*.

I remember reading [this piece](#) by Nathan Robinson, where he compares his own liberal principles saying that colleges shouldn't endorse war-violence-glorifying film "American Sniper" to some conservatives arguing that colleges

shouldn't endorse homosexuality-glorifying book "Fun Home":

It is hypocrisy for liberals to laugh at and criticize the Duke students who have objected to their summer reading book due to its sexual and homosexual themes. They didn't seem to react similarly when students at other universities tried to get screenings of American Sniper cancelled. If you say the Duke students should open their minds and consume things they disagree with, you should say the same thing about the students who boycotted American Sniper. Otherwise, you do not really have a principled belief that people should respect and take in other opinions, you just believe they should respect and take in your own opinions. How can you think in one case the students are close-minded and sheltered, but in the other think they are open-minded and tolerant? What principled distinction is there that allows you to condemn one and praise the other, other than believing people who agree with you are better?

He proposes a bunch of potential counterarguments, then shoots each counterargument down by admitting that the other side would have a symmetrical counterargument of their own: for example, he believes that "American Sniper" is worse because it's racist and promoting racism is genuinely dangerous to a free society, but then he admits a conservative could say that "Fun Home" is worse because in their opinion it's homosexuality that's genuinely dangerous to a free society. After three or four levels of this, he ends up concluding that he can't come up with a meta-level fundamental difference, but he's going to fight for his values anyway because they're his. I'm not sure what I think of this conclusion, but my main response to his article is *oh my gosh he gets the thing*, where "the thing" is a hard-to-describe ability to understand that other people are going to go down as many levels to defend their self-consistent values as you will to defend yours. It seems silly when I'm saying it like this, and you should probably just read the article, but I've seen *so many people* who lack this basic mental operation that this immediately endeared him to me. I would argue Nathan Robinson has a piece of theory-of-mind that a lot of other people are missing.

Actually, I was kind of also thinking this with his [most recent post](#), which complains about [a Washington Post article](#). The Post argues that because the Democrats support gun control and protest police, they are becoming the "pro-crime party". I'm not sure whether the Post genuinely believes the Democrats are pro-crime by inclination or are just arguing their policies will lead to more crime in a hyperbolic figurative way, but I've certainly seen sources further right make the "genuinely in favor of crime as a terminal value" argument. And this doesn't seem too different from the leftist sources that say Republicans can't *really* care about the lives of the unborn, they're just "anti-woman" as a terminal value. Both proposals share this idea of not being able to understand that other people have different beliefs than you and that their actions proceed naturally from those beliefs. Instead of saying "I believe gun control would increase crime, but Democrats believe the opposite, and from their different perspective banning guns makes sense," they say "I believe gun control would increase crime, Democrats must believe the same, and therefore their demands for gun control must come from sinister motives."

(compare: "Brayden brought the Skittles bag with him for lunch, so he must enjoy eating pennies." Or: "Daddy is refusing to watch Sesame Street with me, so he must be secretly watching it with someone else he likes better instead.")

Here are some other mental operations which seem to me to rise to the level of developmental milestones:

1. **Ability to distinguish "the things my brain tells me" from "reality"** – maybe this is better phrased as "not immediately trusting my system 1 judgments". This is a big part of cognitive therapy – building the understanding that just because your brain makes assessments like "I will definitely fail at this" or "I'm the worst person in the world" doesn't mean that *you* have to believe them. As [Ozy points out](#), this one can be *easier* for people with serious psychiatric problems who have a lot of experience with their brain's snap assessments being *really* off, as opposed to everyone else who has to piece the insight together from [a bunch of subtle failures](#).

2. **Ability to model other people as having really different mind-designs** from theirs; for example, the person who thinks that someone with depression is just “being lazy” or needs to “snap out of it”. This is one of the most important factors in determining whether I get along with somebody – people who don’t have this insight tend not to respect boundaries/preferences very much simply because they can’t believe they exist, and to simultaneously get angry when other people violate their supposedly-obvious-and-universal boundaries and preferences.
3. **Ability to think probabilistically** and tolerate uncertainty. [My thoughts on this](#) were mostly inspired by another of David Chapman’s posts, which I’m starting to think might not be a coincidence.
4. **Understanding the idea of trade-offs**; things like “the higher the threshold value of this medical test, the more likely we’ll catch real cases but also the more likely we’ll get false positives” or “the lower the burden of proof for people accused of crimes, the more likely we’ll get real criminals but also the more likely we’ll encourage false accusations”. When I hear people discuss these cases in real life, they’re almost never able to maintain this tension and almost always collapse it to their preferred plan having no downside.

Framed like this, both psychotherapy and LW-style rationality aim to teach people some of these extra mental operations. The reactions to both vary from enlightenment to boredom to bafflement depending on whether the listener needs the piece, already has the piece, or just plain lacks the socket that the piece is supposed to snap into.

This would have a funny corollary; the [LW Sequences](#) try to hammer in how different other minds can be from your own in order to develop the skill of thinking about artificial intelligences, but *whether or not AI matters* this might be an unusually effective hack to break a certain type of person out of their egocentrism and teach them how to deal with other humans.

This raises the obvious question of whether there are any basic mental operations I still don’t have, how I would recognize them if there were, and how I would learn them once I recognized them.

An Iron Curtain Has Descended Upon Psychopharmacology

Posted on August 16, 2014 by Scott Alexander



Imagine if a chemist told you offhandedly that the Russians had different chemical elements than we did.

Here in America, we use elements like lithium and silicon and bismuth. We have figured out lots of neat compounds we can make with these elements. We've also figured out useful technological applications. Lithium makes batteries. Silicon makes computer chips. Bismuth makes pretty gifs you can post on Tumblr.



The Russians don't use any of these. They have their own Russian elements on their own Russian periodic table, with long Russian names you can't pronounce. Apparently some of these also have useful technological applications. One of them is a room temperature superconductor. Another improves the efficiency of dirigibles by 500% for some reason.

No one in America seems remotely interested in any of these Russian elements. Many American chemists don't even know they exist, even though each element has its own English-language Wikipedia page. When informed, they just say "Yeah, the Russians have *lots* of stuff," and leave it at that.

American research teams pour millions of dollars into synthesizing novel elements in order to expand their periodic tables and the number of useful compounds they can make. If anyone suggests importing and studying some of the Russian elements, the chemists say "Huh, that never occurred to us, maybe someone else should do it," and go back to spending millions of dollars synthesizing entirely novel atoms.

If a chemist told you this, you would think they were crazy. Science, you would say, is science everywhere. You can't have one set of elements in Russia and another in the US, everyone would work together and compare notes. At the

very least one side would have the common decency to at least steal from the other. No way anything like this could possibly go on.

But as far as I can tell this is exactly the state of modern psychopharmacology.

Consider anxiety. I would *kill* for a good anti-anxiety drug. Right now my choices are pretty limited. Benzodiazepines and barbiturates work great but are addictive and dangerous. SSRIs work okay but need a month to take effect. Neurontin, Vistaril, and Buspar are safe, fast-acting, and totally ineffective. And Lyrica is expensive and off-label. As a result, a lot of my anxious patients tend to stay anxious.

Any textbook, database, or lecture you care to check on anti-anxiety medications will list the ones I just listed above plus a couple of others I'm forgetting.

But if you look the matter up on Wikipedia, [you see](#) all these weird names like mebicarum, afobazole, selank, bromantane, emoxypine, validolum, and picamilon. You can show these names to your psychiatrist and she will have no idea what you're talking about, think you're speaking nonsense syllables. You can show them to the professor of psychopharmacology at a major university and your chances are maybe like 50-50.

These are the Russian anti-anxiety drugs. They seem to have pretty good evidential support. Wikipedia's bromantane article gives a bunch of studies of bromantane in the footnotes, including [a randomized controlled trial](#) in the forbiddingly named *Zh Nevrol Psichiatr Im S S Korsakova*.

And look what else Wikipedia's bromantane article says:

Study results suggest that the combination of psychostimulant and anxiolytic actions in the spectrum of psychotropic activity of bromantane is effective in treating asthenic disorders compared to placebo. It is considered novel having both stimulant and anti-anxiety properties.

Imagine reading about a Russian element on Wikipedia, and at the end there's this paragraph saying "By the way, this element inverts gravity and has to be tied to the ground to prevent it from falling upwards". An anxiolytic stimulant is *really really cool*. But somehow generations of American psychopharmacologists must have read about bromantane and thought "No, I don't think I'll pay any more attention to *that*."

My guess is the reason we can't prescribe bromantane is the same reason we [can't prescribe melatonin](#) and we [can't prescribe fish oil](#) without the charade of calling it LOVAZA™®©. The FDA won't approve a treatment unless some drug company has invested a billion dollars in doing a lot of studies about it. It doesn't count if some foreign scientists already did a bunch of studies. It doesn't count if millions of Russians have been using the drug for decades and are by and large still alive. You've got to have the entire thing analyzed by the FDA and then rejected at the last second without explanation (yes, I have just been reading [Marginal Revolution's review of Innovation Breakdown: How the FDA and Wall Street Cripple Medical Advances](#); I do need to check out the actual book). Absent an extremely strong patent on the drug there's no reason a drug company would want to go forward with all of this. I don't know what the legalities of buying Russian drug rights from Russian companies are, but I expect they're complicated and that pharmaceutical companies have made a reasoned decision not to bother.

Given this situation, it's perfectly reasonable for doctors not to prescribe them. Certainly I don't plan to prescribe any Russian drugs when I get my own practice. Imagine if a patient gets liver failure on one – and remember that people are getting liver failure all the time for random reasons. The patient's family decides to sue and I'm stuck defending my decision in court. "Yes, Your Honor, I admit I told the deceased to buy a medication no other psychiatrist in the state has ever heard of from a sketchy online Russian pharmacy. But in my defense, there was a study supporting its use in *Zh Nevrol Psichiatr Im S S Korsakova*. Which I didn't read, because I don't speak Russian."

Everyone follows their own incentives perfectly, and as a result the system as a whole does something insane. Classic [multipolar trap](#).

Luckily, this hasn't stopped a lively gray market trade in these chemicals, which I totally one hundred percent approve of. Noopept, for example, is a prescription drug in Russia but is sold over-the-counter by online suppliers here. You can even [get some bromantane](#) for two bucks a pill.

Don't worry. I'm sure these people are on the level. How could a site with a background like that possibly be unreliable?

Why Were Early Psychedelicists So Weird?

Posted on April 28, 2016 by Scott Alexander



Epistemic status: very speculative, asserted with only ~30% confidence. On the other hand, even though psychiatrists don't really talk about this it's possible other groups know this all already.

*
**

A few weeks ago I gave a presentation on the history of early psychedelic research. Since I had a tough crowd, I focused on the fascinating biographies of some of the early psychedelicists.

[Timothy Leary](#) was a Harvard professor and former NIMH researcher who made well-regarded contributions to psychotherapy and psychometrics. He started the Harvard Psilocybin Project and several other Harvard-based experiments to test the effects of psychedelics on normal and mentally ill subjects. He was later fired from Harvard and arrested; later he accomplished a spectacular break out of prison and fled to Algeria. During his later life, he wrote books about how the human brain had hidden circuits of consciousness that would allow us to live in space, including a quantum overmind which could control reality and break the speed of light. He eventually fell so deep into madness that he started hanging out with Robert Anton Wilson and participating in Ron Paul fundraisers.

[Richard Alpert](#) was Leary's co-investigator at the Harvard Psilocybin Project. He, too, had all the signs of a promising career, including a psychology PhD from Stanford, a visiting professorship at Berkeley, and a combination academic/clinical position at Stanford. After his work with Leary, he moved to India, changed his name to Baba Ram Dass, and became one of the world's most prominent advocates for bhakti yoga.

[John Lilly](#) was a doctor, a neuroanatomy researcher, and an inventor who helped develop the principle behind many modern neuroprosthetics. He was always very strange, and did a lot of work in human-dolphin communication and SETI even before starting his work with LSD. But in the 1960s, he ran across Richard Alpert, joined in his LSD experiments, and became even stranger. He started writing books with names like "Programming And Metaprogramming The Human Biocomputer", and arguing that benevolent and malevolent aliens were locked in a battle to manipulate Earth's coincidences and with them the future of the human species. He became an expert yogi and claimed to have achieved samadhi, the highest state of union with God.

[Kary Mullis](#) is kind of cheating since he was not technically a psychedelicist. He was a biochemist in the completely unrelated field of bacterial iron transport molecules. But he did try LSD in 1966 back when it was still a legal research chemical. In fact he tried 1000 micrograms of it, one of the biggest doses I've ever heard of someone taking. Like the others, Mullis was a brilliant scientist – he won the Nobel Prize in Chemistry for inventing the polymerase chain reaction. Like the others, Mullis got *really weird* fast. He is a global warming denialist, HIV/AIDS denialist, and ozone hole denialist; on the other hand, he *does* believe in the efficacy of astrology. He also believes he has contacted extraterrestrials in the form of a fluorescent green raccoon, and "founded a business with the intent to sell pieces of jewelry containing the amplified DNA of deceased famous people like Elvis Presley".

I wondered if there might be a selection bias in which psychedelicists I heard about, or that I might be cherry-picking the most unusual examples, so I looked for leading early psychedelics researchers I'd never heard of before and checked how weird *they* were. My sources told me that the two most important early psychedelicists were [Humphry Osmond](#) (who invented the word 'psychedelic' and may have been the first person to experiment with LSD rigorously) and his colleague [John Smythies](#).

Osmond has an impressive early resume: started off as a surgeon, became a psychiatrist, did some well-regarded research into the structure of the human metabolite adrenochrome. And although he did not become fluorescent-alien-raccoon level weird, he can't quite be called normal either. He became one of the founders of orthomolecular psychiatry, a discipline arguing that schizophrenia and other psychiatric diseases can be cured by massive amounts of vitamins – this is currently considered pseudoscience. His publications include the article "Selection of twins for ESP experimentation" in *International Journal of Parapsychology*, and a history of psychedelics [records](#) that "after his mescaline experiment in 1951, Dr. Osmond claimed to have successfully transmitted telepathic information to a fellow researcher, Duncan Blewett, who was also under the influence of mescaline, leading an independent observer to panic at the uncanny event." He seems to have maintained a lifetime interest in parapsychology, Jungian typological analysis, and a field of his own invention called "socio-architecture".

Smythies was a neuropsychiatrist, neuroanatomist, biochemist, EEG researcher, editor of the International Review of Neurobiology, etc, etc, etc (also, a cousin of Richard Dawkins). He is 94 but apparently still alive and going strong and making new neuroanatomical discoveries. He was one of the first people to investigate the pharmacology of psychedelics and helped with Osmond's experiments in the early 1950s. He has also written *The Walls Of Plato's Cave*, a book presenting a new theory of consciousness which "extends our concepts of consciousness and analyses possible geometrical and topological relations between phenomenal space and physical space linked to brane theory in physics" (I kind of wish I was a fly on the wall at his and Dawkins' family reunions).

My point is that the field of early psychedelic research seemed to pretty consistently absorb brilliant scientists, then spit out people who, while still brilliant scientists, also had styles of thought that could be described as extremely original at best and downright crazy at worst.

I think it's important to try to understand why.

First possibility: you had to be kind of weird to begin with in order to be interested in researching psychedelics. On the one hand, this is surely true; on the other, the early psychedelists ended up *really* weird. At least in the early days I'm not sure psychedelics had the reputation for weirdness they now enjoy, and I'm also not sure that we're living in a world where a high enough percent of psychiatrists go off to become gurus in India, that we can just dismiss LSD research as happening to attract that type of person.

Second possibility: I know that almost all of these researchers (I'm not sure about Smythies) used psychedelics themselves. Psychedelic use is a sufficiently interesting experience that I can see why it might expand one's interest in the study of consciousness and the universe. Perhaps this is especially true if you're one of the first people to use it, and you don't have the social setting of "Oh, yeah, this is that drug that makes you have really weird experiences about consciousness for a while". If you're not aware that psychedelic hallucinations are a thing that happens, you might have to interpret your experience in more traditional terms like divine revelation. Under this theory, these pioneers had to become kind of weird to learn enough for the rest of us to use these substances safely. But why would that make John Lilly obsessed with aliens? Why would it turn Timothy Leary into a space colonization advocate and Ron Paul supporter?

The third possibility is the one that really intrigues me. [A 2011 study](#) found that a single dose of psilocybin *could permanently increase the personality dimension of Openness To Experience*. I'm emphasizing that because personality is otherwise pretty stable after adulthood; *nothing* should be able to do this. But magic mushrooms apparently have this effect, and not subtly either; participants who had a mystical experience on psilocybin had Openness increase *up to half a standard deviation* compared to placebo, and *the change was stable sixteen months later*. This is *really scary*. I mean, I *like* Openness To Experience, but something that can produce large, permanent personality changes is so far beyond anything else we have in psychiatry that it's kind of terrifying.

(related: 1972 study finds LSD may cause [permanent increase in hypnotic susceptibility](#), which [other sources](#) have

linked to being “fantasy prone” and “creative”)

And that’s *one dose*. These researchers were taking psychedelics pretty constantly for years, and probably experimented with the sort of doses you couldn’t get away with giving research subjects. What would you expect to happen to *their Openness To Experience*? How many standard deviations do you think *it* went up?

It seems possible to me that psychedelics have a direct pharmacological effect on personality that causes people to be more open to unusual ideas. I know this is going against most of the latest research, which says [psychedelics have no long-term negative mental health effects](#) and do not cause psychosis. But there’s a difference between being schizophrenic, and being the sort of guy who is still a leading neuroanatomist but also writes books about the geometric relationships between consciousness and the space-time continuum.

I’m not sure anyone has ever done studies to rule out the theory that psychedelics just plain make people *weird*. Indeed, such studies would be very difficult, given that weird people with very high Openness To Experience are more likely to use psychedelics. This problem would even prevent common sense detection of the phenomenon – even if we noticed that frequent psychedelic users were really weird, we would attribute it to selection effects and forget about it.

In this situation, the early psychedelicists could be a natural experiment giving us data we can’t get any other way. Here are relatively sober scientists who took psychedelics for reasons other than being weird hippies already. Their fate provides signal through the noise which is the general psychedelic-using population.

I think this is only medium-risk; the explanation that weird people gravitate toward psychedelics, even in the sciences, is a strong one. But it’s sufficient that I am hesitant to repeat the common view that psychedelics are not at all dangerous, or that they have no permanent side effects. There seems to me at least a moderate chance that they will make you more interesting without your consent – whether that is a good or a bad thing depends on exactly how interesting you want to be.

Universal Love, Said The Cactus Person

Posted on April 21, 2015 by Scott Alexander



"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

"Right," I said. "I'm absolutely in favor of both those things. But before we go any further, could you tell me the two prime factors of
1,522,605,027,922,533,360,535,618,378,132,637,429,718,068,114,961,380,688,657,908,494,580,122,963,258,952,89

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

The sea was made of strontium; the beach was made of rye. Above my head, a watery sun shone in an oily sky. A thousand stars of sertraline whirled round quetiapine moons, and the sand sizzled sharp like cooking oil that hissed and sang and threatened to boil the octahedral dunes.

"Okay," I said. "Fine. Let me tell you where I'm coming from. I was reading [Scott McGreal's blog](#), which has some [good articles](#) about so-called DMT entities, and mentions how they seem so real that users of the drug insist they've made contact with actual superhuman beings and not just psychedelic hallucinations. You know, [the usual](#) Terence McKenna stuff. But in [one](#) of them he mentions a paper by Marko Rodriguez called [A Methodology For Studying Various Interpretations of the N,N-dimethyltryptamine-Induced Alternate Reality](#), which suggested among other things that you could prove DMT entities were real by taking the drug and then asking the entities you meet to factor large numbers which you were sure you couldn't factor yourself. So to that end, could you do me a big favor and tell me the factors of
1,522,605,027,922,533,360,535,618,378,132,637,429,718,068,114,961,380,688,657,908,494,580,122,963,258,952,89

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

The sea turned hot and geysers shot up from the floor below. First one of wine, then one of brine, then one more yet of turpentine, and we three stared at the show.

"I was afraid you might say that. Is there anyone more, uh, *verbal* here whom I could talk to?"

"Universal love," said the cactus person.

At the sound of that, the big green bat started rotating in place. On its other side was a bigger greener bat, with a ancient, wrinkled face.

"*Not splitting numbers / but joining Mind*," it said.

Not facts or factors or factories / but contact with the abstract attractor that brings you back to me
Not to seek / but to find"

"I don't follow," I said.

*"Not to follow / but to jump forth into the deep
Not to grind or to bind or to seek only to find / but to accept
Not to be kept / but to wake from sleep"*

The bat continued to rotate, until the first side I had seen swung back into view.

"Okay," I said. "I'm going to hazard a guess as to what you're talking about, and you tell me if I'm right. You're saying that, like, all my Western logocentric stuff about factoring numbers in order to find out the objective truth about this realm is missing the point, and I should be trying to do some kind of spiritual thing involving radical acceptance and enlightenment and such. Is that kind of on the mark?"

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

"Frick," I said. "Well, okay, let me continue." The bat was still rotating, and I kind of hoped that when the side with the creepy wrinkled face came into view it might give me some better conversation. "I'm all about the spiritual stuff. I wouldn't be here if I weren't deeply interested in the spiritual stuff. This isn't about money or fame or anything. I want to advance psychedelic research. If you can factor that number, then it will convince people back in the real – back in my world that this place is for real and important. Then lots of people will take DMT and flock here and listen to what you guys have to say about enlightenment and universal love, and make more sense of it than I can alone, and in the end we'll have more universal love, and... what was the other thing?"

"Transcendent joy," said the big green bat.

"Right," I said. "We'll have more transcendent joy if you help me out and factor the number than if you just sit there being spiritual and enigmatic."

*"Lovers do not love to increase the amount of love in the world / But for the mind that thrills
And the face of the beloved, which the whole heart fills / the heart and the art never apart, ever unfurled
And John Stuart is one of / the dark satanic mills"*

"I take it you're not consequentialists," I said. "You know that's really weird, right. Like, not just 'great big green bat with two faces and sapient cactus-man' weird, but like *really* weird. You talk about wanting this spiritual enlightenment stuff, but you're not going to take actions that are going to increase the amount of spiritual enlightenment? You've got to understand, this is like a bigger gulf for me than normal human versus ineffable DMT entity. You can have crazy goals, I expect you to have crazy goals, but what you're saying now is that you don't pursue any goals at all, you can't be modeled as having desires. Why would you *do* that?"

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

"Now you see here," I said. "Everyone in this conversation is in favor of universal love and transcendent joy. But I've seen the way this works. Some college student gets his hands on some DMT, visits here, you guys tell him about universal love and transcendent joy, he wakes up, says that his life has been changed, suddenly he truly understands what really matters. But it never lasts. The next day he's got to get up and go to work and so on, and the universal love lasts about five minutes until his boss starts yelling at him for writing his report in the wrong font, and before you

know it twenty years later he's some slimy lawyer who's joking at a slimy lawyer party about the one time when he was in college and took some DMT and spent a whole week raving about transcendent joy, and all the other slimy lawyers laugh, and he laughs with them, and so much for whatever spiritual awakening you and your colleagues in LSD and peyote are trying to kindle in humanity. And if I accept your message of universal love and transcendent joy right now, that's exactly what's going to happen to me, and meanwhile human civilization is going to keep being stuck in greed and ignorance and misery. So how about you shut up about universal love and you factor my number for me so we can start figuring out a battle plan for giving humanity a *real* spiritual revolution?"

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

A meteorite of pure delight struck the sea without a sound. The force of the blast went rattling past the bat and the beach, disturbing each, then made its way to a nearby bay of upside-down trees with their roots in the breeze and their branches underground.

"I demand a better answer than that," I demanded.

The other side of the bat spun into view.

*"Chaos never comes from the Ministry of Chaos / nor void from the Ministry of Void
Time will decay us but time can be left blank / destroyed
With each Planck moment ever fit / to be eternally enjoyed"*

"You're making this basic mistake," I told the big green bat. "I honestly believe that there's a perspective from which Time doesn't matter, where a single moment of recognition is equivalent to eternal recognition. The problem is, if you only have that perspective for a moment, then all the rest of the time, you're sufficiently stuck in Time to honestly believe you're stuck in Time. It's like that song about the hole in the bucket – if the hole in the bucket were fixed, you would have the materials needed to fix the hole in the bucket. But since it isn't, you don't. Likewise, if I understood the illusoriness... illusionality... whatever, of time, then I wouldn't care that I only understood it for a single instant. But since I don't, I don't. Without a solution to the time-limitedness of enlightenment that works from *within* the temporal perspective, how can you consider it solved at all?"

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

The watery sun began to run and it fell on the ground as rain. It became a dew that soaked us through, and as the cold seemed to worsen the cactus person hugged himself to stay warm but his spines pierced his form and he howled in a fit of pain.

"You know," I said, "sometimes I think the [kvithion sumurhe](#) had the right of it. The world is an interference pattern between colliding waves of Truth and Beauty, and either one of them pure from the source and undiluted by the other will be fatal. I think you guys and some of the other psychedelics might be pure Beauty, or at least much closer to the source than people were meant to go. I think you can't even understand reason, I think you're constitutionally opposed to reason, and that the only way we're ever going to get something that combines your wisdom and love and joy with reason is after we immanentize the eschaton and launch civilization into some perfected postmessianic era where the purpose of the world is fully complete. And that as much as I hate to say it, there's no short-circuiting the process."

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

"I'm dissing you, you know. I'm saying you guys are so intoxicated on spiritual wisdom that you couldn't think straight if your life depended on it; that your random interventions in our world and our minds look like the purposeless acts of a drunken madman because that's basically more or less what they are. I'm saying if you had like five IQ points between the two of you, you could tap into your cosmic consciousness or whatever to factor a number that would do more for your cause than all your centuries of enigmatic dreams and unasked-for revelations combined, and you ARE TOO DUMB TO DO IT EVEN WHEN I BASICALLY HOLD YOUR HAND THE WHOLE WAY. Your spine. Your wing. Whatever."

"Universal love," said the cactus person.

"Transcendent joy," said the big green bat.

"Fuck you," said I.

I saw the big green bat bat a green big eye. Suddenly I knew I had gone too far. The big green bat started to turn around what was neither its x, y, or z axis, slowly rotating to reveal what was undoubtedly the biggest, greenest bat that I had ever seen, a bat bigger and greener than which it was impossible to conceive. And the bat said to me:

"Sir. Imagine you are in the driver's seat of a car. You have been sitting there so long that you have forgotten that it is the seat of a car, forgotten how to get out of the seat, forgotten the existence of your own legs, indeed forgotten that you are a being at all separate from the car. You control the car with skill and precision, driving it wherever you wish to go, manipulating the headlights and the windshield wipers and the stereo and the air conditioning, and you pronounce yourself a great master. But there are paths you cannot travel, because there are no roads to them, and you long to run through the forest, or swim in the river, or climb the high mountains. A line of prophets who have come before you tell you that the secret to these forbidden mysteries is an ancient and terrible skill called GETTING OUT OF THE CAR, and you resolve to learn this skill. You try every button on the dashboard, but none of them is the button for GETTING OUT OF THE CAR. You drive all of the highways and byways of the earth, but you cannot reach GETTING OUT OF THE CAR, for it is not a place on a highway. The prophets tell you GETTING OUT OF THE CAR is something fundamentally different than anything you have done thus far, but to you this means ever sillier extremities: driving backwards, driving with the headlights on in the glare of noon, driving into ditches on purpose, but none of these reveal the secret of GETTING OUT OF THE CAR. The prophets tell you it is easy; indeed, it is the easiest thing you have ever done. You have traveled the Pan-American Highway from the boreal pole to the Darien Gap, you have crossed Route 66 in the dead heat of summer, you have outrun cop cars at 160 mph and survived, and GETTING OUT OF THE CAR is easier than any of them, the easiest thing you can imagine, closer to you than the veins in your head, but still the secret is obscure to you."

A herd of bison came into listen, and voles and squirrels and ermine and great tusked deer gathered round to hear as the bat continued his sermon.

"And finally you drive to the top of the highest peak and you find a sage, and you ask him what series of buttons on the dashboard you have to press to get out of the car. And he tells you that it's not about pressing buttons on the dashboard and you just need to GET OUT OF THE CAR. And you say okay, fine, but what series of buttons will *lead to* you getting out of the car, and he says no, really, you need to stop thinking about dashboard buttons and GET OUT OF THE CAR. And you tell him maybe if the sage helps you change your oil or rotates your tires or something then it will improve your driving to the point where getting out of the car will be a cinch after that, and he tells you it has nothing to do with how rotated your tires are and you just need to GET OUT OF THE CAR, and so you call him a moron and

drive away."

"Universal love," said the cactus person.

"So that metaphor is *totally unfair*," I said, "and a better metaphor would be if every time someone got out of the car, five minutes later they found themselves back in the car, and I ask the sage for driving directions to a laboratory where they are studying that problem, and..."

"You only believe that because it's written on the windshield," said the big green bat. "And you think the windshield is identical to reality because you won't GET OUT OF THE CAR."

"Fine," I said. "Then I can't get out of the car. I want to get out of the car. But I need help. And the first step to getting help is for you to factor my number. You seem like a reasonable person. Bat. Freaky DMT entity. Whatever. Please. I promise you, this is the right thing to do. Just factor the number."

"And I promise you," said the big green bat. "You don't need to factor the number. You just need to GET OUT OF THE CAR."

"I can't get out of the car until you factor the number."

"I won't factor the number until you get out of the car."

"Please, I'm begging you, factor the number!"

"Yes, well, I'm begging you, please get out of the car!"

"FOR THE LOVE OF GOD JUST FACTOR THE FUCKING NUMBER!"

"FOR THE LOVE OF GOD JUST GET OUT OF THE FUCKING CAR!"

"FACTOR THE FUCKING NUMBER!"

"GET OUT OF THE FUCKING CAR!"

"Universal love," said the cactus person.

Then tree and beast all fled due east and the moon and stars shot south. And the bat rose up and the sea was a cup and the earth was a screen green as clozapine and the sky a voracious mouth. And the mouth opened wide and the earth was skied and the sea fell in with an awful din and the trees were moons and the sand in the dunes was a blazing comet and...

I vomited, hard, all over my bed. It happens every time I take DMT, sooner or later; I've got a weak stomach and I'm not sure the stuff I get is totally pure. I crawled just far enough out of bed to flip a light switch on, then collapsed back onto the soiled covers. The clock on the wall read 11:55, meaning I'd been out about an hour and a half. I briefly considered taking some more ayahuasca and heading right back there, but the chances of getting anything more out of the big green bat, let alone the cactus person, seemed small enough to fit in a thimble. I drifted off into a fitful sleep.

Behind the veil, across the infinite abyss, beyond the ice, beyond death, the dew rose from the soaked ground and coalesced into a great drop, which floated up into an oily sky and became a watery sun. The cactus person was counting on his spines.

"Hey," the cactus person finally said, "just out of curiosity, was the answer
37,975,227,936,943,673,922,808,872,755,445,627,854,565,536,638,199 times
40,094,690,950,920,881,030,683,735,292,761,468,389,214,899,724,061?"

"Yeah," said the big green bat. "That's what I got too."

Did A Melatonin Patent Inspire Current Dose Confusion?

Posted on July 12, 2018 by Scott Alexander



Yesterday [I wrote about melatonin](#), mentioning that most drugstore melatonin supplements were 10x or more the recommended dose. A commenter on Facebook pointed me to an interesting explanation of why.

Dr. Richard Wurtman, an MIT scientist who helped discover melatonin's role in the body and pioneer its use as a sleep aid, [writes](#):

MIT was so excited about our research team's melatonin-sleep connection discovery that they decided to patent the use of reasonable doses of melatonin—up to 1 mg—for promoting sleep.

But they made a big mistake. They assumed that the FDA would want to regulate the hormone and its use as a sleep therapy. They also thought the FDA wouldn't allow companies to sell melatonin in doses 3-times, 10-times, even 15-times more than what's necessary to promote sound sleep.

Much to MIT's surprise, however, the FDA took a pass on melatonin. At that time, the FDA was focusing on other issues, like nicotine addiction, and they may have felt they had bigger fish to fry.

Also, the FDA knew that the research on melatonin showed it to be non-toxic, even at extremely high doses, so they probably weren't too worried about how consumers might use it. In the end, and as a way of getting melatonin on to the market, the FDA chose to label it a dietary supplement, which does not require FDA regulation. Clearly, this was wrong because melatonin is a hormone, not a dietary supplement.

Quickly, supplement manufacturers saw the huge potential in selling melatonin to promote good sleep. After all, millions of Americans struggled to get to sleep and stay asleep, and were desperate for safe alternatives to anti-anxiety medicines and sleeping pills that rarely worked well and came with plenty of side effects.

Also, manufacturers must have realized that they could avoid paying royalties to MIT for melatonin doses over the 1 mg measure. So, they produced doses of 3 mg, 5 mg, 10 mg and more! Their thinking—like so much else in our American society—was likely, “bigger is better!” But, they couldn’t be more wrong.

So he's saying that... in order to get around a patent on using the correct dose of melatonin... supplement manufacturers... used the wrong dose of melatonin? I enjoy collecting stories of [all the crazy perversities](#) created by our current pharmaceutical system, but this one really takes the cake.

Assuming it's true, that is. Commenter Rodrigo [brings up](#) some reasons to be suspicious:

1. Who would patent a drug only up to a certain dose? Isn't this really dumb?
2. To avoid the patent on the correct dose, drugstores just have to sell more than 1 mg – for example, 2 mg. But they actually sell up to 10 mg.

To these I would add:

1. Lots of supplements are very high dose. When I Google Vitamin C, the [first product](#) that comes up advertises that it has 1111% of the recommended daily allowance, which seems better optimized for numerological purposes than medical ones.
2. A few companies do sell melatonin at the right dose range, and MIT hasn't sued them yet.

Normally I would find these considerations pretty persuasive, but I feel like the guy who discovered melatonin and ran a pharmaceutical company for a while knows more about the history of melatonin and pharmaceutical regulations than I do.

From [last week](#):

This kind of thing is the endless drudgery of rationality training... questions like "How much should you discount a compelling-sounding theory based on the bias of its inventor?" And "How much does someone being a famous expert count in their favor?" And "How concerned should we be if a theory seems to violate efficient market assumptions?" And "How do we balance arguments based on what rationally has to be true, vs. someone's empirical but fallible data sets?"

Here I'm just really skeptical of the MIT patent story. Wurtman seems to admit that "bigger is better" played a role. Maybe the patent thing was a very small issue, around the beginning of melatonin sales, and was soon forgotten – but the tradition of expecting melatonin to be very high dose stuck around forever, mostly for other reasons?

EDIT: Commenters, including a patent lawyer, have filled in the rest of the story. Because melatonin is a natural hormone and not an invention, patents can only cover specific uses of it. The MIT patent covered the proper way to use it for sleep; a broader patent might not have been granted. The patent probably guided supplement companies, but expired about five years ago. It's now legal to produce melatonin 0.3 mg pills, but people are so used to higher doses that few people do.

It's Bayes All The Way Up

Posted on September 12, 2016 by Scott Alexander



Epistemic status: Very speculative. I am not a neuroscientist and apologize for any misinterpretation of the papers involved. Thanks to the people who posted these papers in [r/slatestarcode](#). See also [Mysticism and Pattern-Matching](#) and [Bayes For Schizophrenics](#).

I

Bayes' Theorem is an equation for calculating certain kinds of conditional probabilities. For something so obscure, it's attracted a surprisingly wide fanbase, including [doctors](#), [environmental scientists](#), [economists](#), [bodybuilders](#), [fendwellers](#), and [international smugglers](#). Eventually the hype reached the point where there was both a [Bayesian cabaret](#) and a [Bayesian choir](#), popular books using Bayes' Theorem to prove both the [existence](#) and the [nonexistence](#) of God, and even [Bayesian dating advice](#). Eventually everyone agreed to dial down their exuberance a little, and accept that Bayes' Theorem might not literally explain *absolutely* everything.

So – did you know that the neurotransmitters in the brain might represent different terms in Bayes' Theorem?

First things first: Bayes' Theorem is a mathematical framework for integrating new evidence with prior beliefs. For example, suppose you're sitting in your quiet suburban home and you hear something that sounds like a lion roaring. You have some prior beliefs that lions are unlikely to be near your house, so you figure that it's probably not a lion. Probably it's some weird machine of your neighbor's that just happens to sound like a lion, or some kids pranking you by playing lion noises, or something. You end up believing that there's probably no lion nearby, but you do have a slightly higher probability of there being a lion nearby than you had before you heard the roaring noise. Bayes' Theorem is just this kind of reasoning converted to math. You can find the long version [here](#).

This is what the brain does too: integrate new evidence with prior beliefs. Here are some examples I've used on this blog before:

THE CAT

I
LOVE
PARIS IN THE
THE SPRINGTIME



All three of these are examples of top-down processing. Bottom-up processing is when you build perceptions into a model of the world. Top-down processing is when you let your models of the world influence your perceptions. In the first image, you view the center letter of the word as an H and the second as an A, even though they're the same character; your model of the world tells you that THE CAT is more likely than TAE CHT. In the second image, you read "PARIS IN THE SPRINGTIME", skimming over the duplication of the word "the"; your model of the world tells you that the phrase should probably only have one "the" in it (just as you've probably skimmed over it three times I've duplicated "the" in this paragraph alone!). The third image might look meaningless until you realize it's a cow's head; once you see the cow's head your model of the world informs your perception and it's almost impossible to see it as anything else.

(Teh fcatt taht you can siltl raed wrods wtih all the itroneir ltretrs rgraneanrd is ahonter empxlae of top-dwon pssirocneg mkinag nsioy btotom-up dtaa sanp itno pacle)

But top-down processing is much more omnipresent than even these examples would suggest. Even something as simple as looking out the window and seeing a tree requires top-down processing; it may be too dark or foggy to see the tree one hundred percent clearly, the exact pattern of light and darkness on the tree might be something you've never seen before – but because you know what trees are and expect them to be around, the image "snaps" into the schema "tree" and you see a tree there. As usual, this process is most obvious when it goes wrong; for example, when random patterns on a wall or ceiling "snap" into the image of a face, or when the whistling of the wind "snaps" into a voice calling your name.

Most of the things you perceive when awake are generated from very limited input – by the same machinery that generates dreams with no input

— [Void Of Space \(@VoidOfSpace\), September 2, 2016](#)

[Corlett, Frith & Fletcher \(2009\)](#) (henceforth CFF) expand on this idea and speculate on the biochemical substrates of each part of the process. They view perception as a "handshake" between top-down and bottom-up processing. Top-down models predict what we're going to see, bottom-up models perceive the real world, then they meet in the middle and compare notes to calculate a prediction error. When the prediction error is low enough, it gets smoothed over into

a consensus view of reality. When the prediction error is too high, it registers as salience/surprise, and we focus our attention on the stimulus involved to try to reconcile the models. If it turns out that bottom-up was right and top-down was wrong, then we adjust our priors (ie the models used by the top-down systems) and so learning occurs.

In their model, bottom-up sensory processing involves glutamate via the AMPA receptor, and top-down sensory processing involves glutamate via the NMDA receptor. Dopamine codes for prediction error, and seem to represent the level of certainty or the “confidence interval” of a given prediction or perception. Serotonin, acetylcholine, and the others seem to modulate these systems, where “modulate” is a generic neuroscientist weasel word. They provide a lot of neurological and radiologic evidence for these correspondences, for which I highly recommend reading the paper but which I’m not going to get into here. What I found interesting was their attempts to match this system to known pharmacological and psychological processes.

CFF discuss a couple of possible disruptions of their system. Consider *increased* AMPA signaling combined with *decreased* NMDA signaling. Bottom-up processing would become more powerful, unrestrained by top-down models. The world would seem to become “noisier”, as sensory inputs took on a life of their own and failed to snap into existing categories. In extreme cases, the “handshake” between exuberant bottom-up processes and overly timid top-down processes would fail completely, which would take the form of the sudden assignment of salience to a random stimulus.

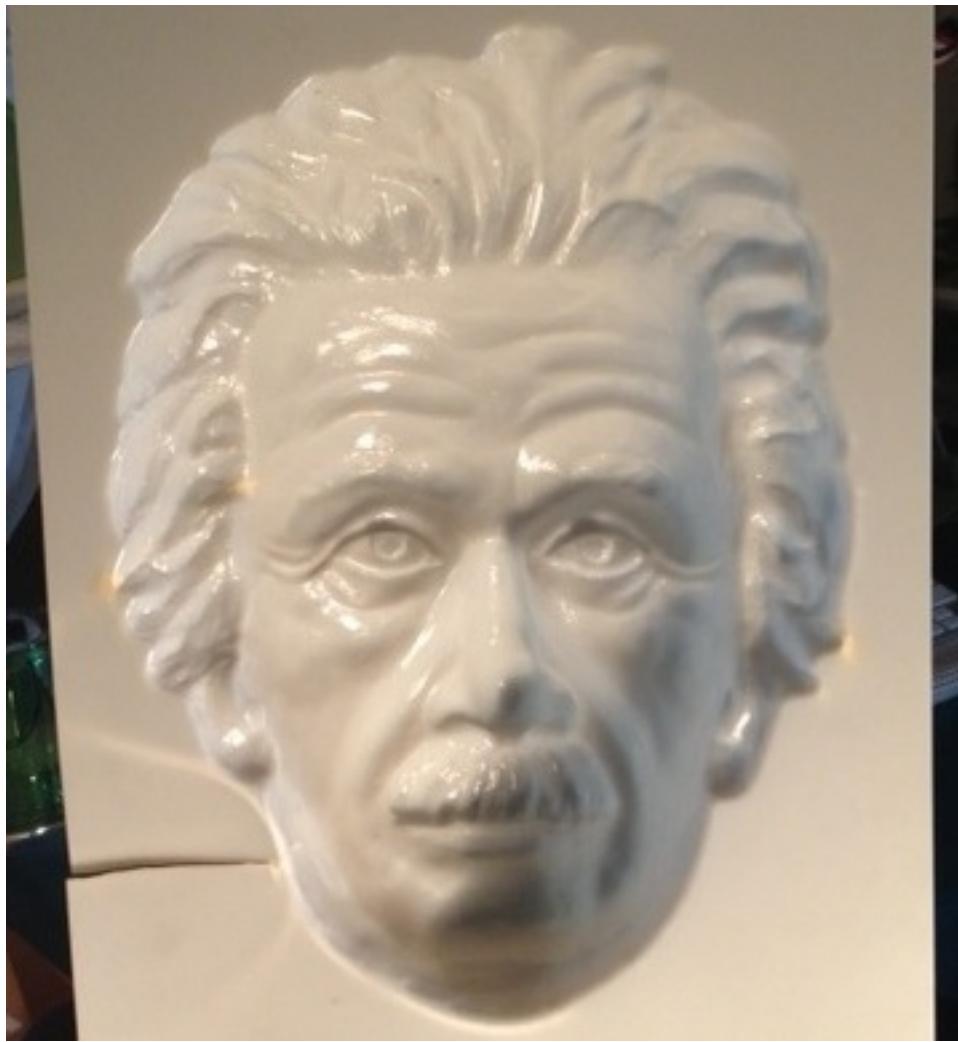
Schizophrenics are famous for “delusions of reference”, where they think a random object or phrase is deeply important for reasons they have trouble explaining. Wikipedia gives as examples:

- A feeling that people on television or radio are talking about or talking directly to them
- Believing that headlines or stories in newspapers are written especially for them
- Seeing objects or events as being set up deliberately to convey a special or particular meaning to themselves
- Thinking ‘that the slightest careless movement on the part of another person had great personal meaning... increased significance’

In CFF, these are perceptual handshake failures; even though “there’s a story about the economy in today’s newspaper” should be perfectly predictable, noisy AMPA signaling registers it as an extreme prediction failure, and it fails its perceptual handshake with overly-weak priors. Then it gets flagged as shocking and deeply important. If you’re unlucky enough to have your brain flag a random newspaper article as shocking and deeply important, maybe phenomenologically that feels like it’s a secret message for you.

And this pattern – increased AMPA signaling combined with decreased NMDA signaling – is pretty much the effect profile of the drug ketamine, and ketamine does cause a paranoid psychosis mixed with delusions of reference.

Organic psychosis like schizophrenia might involve a similar process. There’s a test called the binocular depth inversion illusion, which looks like this:



([source](#))

The mask in the picture is concave, ie the nose is furthest away from the camera. But most viewers interpret it as convex, with the nose closest to the camera. This makes sense in terms of Bayesian perception; we see right-side-in faces a whole lot more often than inside-out faces.

Schizophrenics (and people stoned on marijuana!) are more likely to properly identify the face as concave than everyone else. In CFF's system, something about schizophrenia and marijuana messes with NMDA, impairs priors, and reduces the power of top-down processing. This predicts that schizophrenics and potheads would both have paranoia and delusions of reference, which seems about right.

Consider a slightly different distortion: *increased AMPA signaling combined with increased NMDA signaling*. You've still got a lot of sensory noise. But you've also got stronger priors to try to make sense of them. CFF argue these are the perfect conditions to create hallucinations. The increase in sensory noise means there's a lot of data to be explained; the increased top-down pattern-matching means that the brain is very keen to fit all of it into some grand narrative. The result is vivid, convincing hallucinations of things that are totally not there at all.

LSD is mostly serotonergic, but most things that happen in the brain bottom out in glutamate eventually, and LSD bottoms out in exactly the pattern of increased AMPA and increased NMDA that we would expect to produce hallucinations. CFF don't mention this, but I would also like to add my theory of [pattern-matching based mysticism](#). Make the top-down prior-using NMDA system strong enough, and the entire world collapses into a single narrative, a divine grand plan in which everything makes sense and you understand all of it. This is also something I associate with

LSD.

If dopamine represents a confidence interval, then increased dopaminergic signaling should mean narrowed confidence intervals and increased certainty. Perceptually, this would correspond to increased sensory acuity. More abstractly, it might increase “self-confidence” as usually described. Amphetamines, which act as dopamine agonists, do both. Amphetamine users report increased visual acuity (weirdly, they also report blurred vision sometimes; I don’t understand exactly what’s going on here). They also create an elevated mood and grandiose delusions, making users more sure of themselves and making them feel like they can do anything.

(something I remain confused about: elevated mood and grandiose delusions are also typical of bipolar mania. People on amphetamines and other dopamine agonists act pretty much exactly like manic people. Antidopaminergic drugs like olanzapine are very effective acute antimanic drugs. But people don’t generally think of mania as primarily dopaminergic. Why not?)

CFF end their paper with a discussion of sensory deprivation. If perception is a handshake between bottom-up sense-data and top-down priors, what happens when we turn the sense-data off entirely? Psychologists note that most people go a little crazy when placed in total sensory deprivation, but that schizophrenics actually seem to do *better* under sense-deprivation conditions. Why?

The brain filters sense-data to adjust for ambient conditions. For example, when it’s very dark, your eyes gradually adjust until you can see by whatever light is present. When it’s perfectly silent, you can hear the proverbial pin drop. In a state of total sensory deprivation, any attempt to adjust to a threshold where you can detect the nonexistent signal is actually just going to bring you down below the point where you’re picking up noise. As with LSD, when there’s too much noise the top-down systems do their best to impose structure on it, leading to hallucinations; when they fail, you get delusions. If schizophrenics have inherently noisy perceptual systems, such that all perception comes with noise the same way a bad microphone gives off bursts of static whenever anyone tries to speak into it, then their brains will actually become *less* noisy as sense-data disappears.

(this might be a good time to remember that [no congenitally blind people ever develop schizophrenia](#) and no one knows why)

II

Lawson, Rees, and Friston (2014) offer [a Bayesian link to autism](#).

(there are probably a lot of links between Bayesians and autism, but this is the only one that needs a journal article)

They argue that autism is a form of *aberrant precision*. That is, confidence intervals are too low; bottom-up sense-data cannot handshake with top-down models unless they’re almost-exactly the same. Since they rarely are, top-down models lose their ability to “smooth over” bottom-up information. The world is full of random noise that fails to cohere into any more general plan.

Right now I’m sitting in a room writing on a computer. A white noise machine produces white noise. A fluorescent lamp flickers overhead. My body is doing all sorts of body stuff like digesting food and pumping blood. There are a few things I need to concentrate on: this essay I’m writing, my pager if it goes off, any sorts of sudden dramatic pains in my body that might indicate a life-threatening illness. But I don’t need to worry about the feeling of my back against the back of the chair, or the occasional flickers of the fluorescent light, or the feeling of my shirt on my skin.

A well-functioning perceptual system gates out those things I don’t need to worry about. Since my shirt always feels

more or less similar on my skin, my top-down model learns to predict that feeling. When the top-down model predicts the shirt on my skin, and my bottom-up sensation reports the shirt on my skin, they handshake and agree that all is well. Even if a slight change in posture makes a different part of my shirt brush against my skin than usual, the confidence intervals are wide: it is still an instance of the class “shirt on skin”, it “snaps” into my shirt-on-skin schema, and the perceptual handshake goes off successfully, and all remains well. If something dramatic happens – for example my pager starts beeping really loudly – then my top-down model, which has thus far predicted silence – is rudely surprised by the sudden burst of noise. The perceptual handshake fails, and I am startled, upset, and instantly stop writing my essay as I try to figure out what to do next (hopefully answer my pager). The system works.

The autistic version works differently. The top-down model tries to predict the feeling of the shirt on my skin, but tiny changes in the position of the shirt change the feeling somewhat; bottom-up data does not *quite* match top-down prediction. In a neurotypical with wide confidence intervals, the brain would shrug off such a tiny difference, declare it good enough for government work, and (correctly) ignore it. In an autistic person, the confidence intervals are very narrow; the top-down systems expect the feeling of shirt-on-skin, but the bottom-up systems report a *slightly different* feeling of shirt-on-skin. These fail to snap together, the perceptual handshake fails, and the brain flags it as important; the autistic person is startled, upset, and feels like stopping what they’re doing in order to attend to it.

(in fact, I think the paper might be claiming that “attention” just means a localized narrowing of confidence intervals in a certain direction; for example, if I pay attention to the feeling of my shirt on my skin, then I *can* feel every little fold and micromovement. This seems like an important point with a lot of implications.)

Such handshake failures match some of the sensory symptoms of autism pretty well. Autistic people dislike environments that are (literally or metaphorically) noisy. Small sensory imperfections bother them. They literally get annoyed by scratchy clothing. They tend to seek routine, make sure everything is maximally predictable, and act as if even tiny deviations from normal are worthy of alarm.

They also stim. LRF interpret stimming as an attempt to control sensory predictive environment. If you’re moving your arms in a rhythmic motion, the overwhelming majority of sensory input from your arm is from that rhythmic motion; tiny deviations get lost in the larger signal, the same way a firefly would disappear when seen against the blaze of a searchlight. The rhythmic signal which you yourself are creating and keeping maximally rhythmic is the most predictable thing possible. Even something like head-banging serves to create extremely strong sensory data – sensory data whose production the head-banger is themselves in complete control of. If the brain is in some sense minimizing predictive error, and there’s no reasonable way to minimize prediction error because your predictive system is messed up and registering *everything* as a dangerous error – then sometimes you have to take things into your own hands, bang your head against a metal wall, and say “I totally predicted all that pain”.

(the paper doesn’t mention this, but it wouldn’t surprise me if weighted blankets work the same way. A bunch of weights placed on top of you will predictably stay there; if they’re heavy enough this is one of the strongest sensory signals you’re receiving and it might “raise your average” in terms of having low predictive error)

What about all the non-sensory-gating-related symptoms of autism? LRF think that autistic people dislike social interaction because it’s “the greatest uncertainty”; other people are the hardest-to-predict things we encounter. Neurotypical people are able to smooth social interaction into general categories: this person seems friendly, that person probably doesn’t like me. Autistic people get the same bottom-up data: an eye-twitch here, a weird half-smile there – but it never snaps into recognizable models; it just stays weird uninterpretable clues. So:

This provides a simple explanation for the pronounced social-communication difficulties in autism; given that other agents are arguably the most difficult things to predict. In the complex world of social interactions, the many-to-one mappings between causes and sensory input are dramatically increased and difficult to learn; especially if one cannot contextualize the prediction errors that drive that learning.

They don't really address differences between autists and neurotypicals in terms of personality or skills. But a lot of people have come up with stories about how autistic people are better at tasks that require a lot of precision and less good at tasks that [require central coherence](#), which seems like sort of what this theory would predict.

LRF ends by discussing biochemical bases. They agree with CFF that top-down processing is probably related to NMDA receptors, and so suspect this is damaged in autism. Transgenic mice who lack an important NMDA receptor component seem to behave kind of like autistic humans, which they take as support for their model – although obviously a lot more research is needed. They agree that acetylcholine “modulates” all of this and suggest it might be a promising pathway for future research. They agree with CFF that dopamine may represent precision/confidence, but despite their whole spiel being that precision/confidence is messed up in autism, they don't have much to say about dopamine except that it probably modulates something, just like everything else.

III

All of this is fascinating and elegant. But is it elegant *enough*?

I notice that I am confused about the relative role of NMDA and AMPA in producing hallucinations and delusions. CFF say that enhanced NMDA signaling results in hallucinations as the brain tries to add excess order to experience and “overfits” the visual data. Fine. So maybe you get a tiny bit of visual noise and think you're seeing the Devil. But shouldn't NMDA and top-down processing also be the system that tells you there is a high prior against the Devil being in any particular visual region?

Also, once psychotics develop a delusion, that delusion usually sticks around. It might be that a stray word in a newspaper makes someone think that the FBI is after them, but once they think the FBI is after them, they fit everything into this new paradigm – for example, they might think their psychiatrist is an FBI agent sent to poison them. This sounds a lot like a new, very strong prior! Their doctor presumably isn't doing much that seems FBI-agent-ish, but because they're working off a narrative of the FBI coming to get them, they fit everything, including their doctor, into that story. But if psychosis is a case of attenuated priors, why should that be?

(maybe they would answer that because psychotic people also have increased dopamine, they believe in the FBI with absolute certainty? But then how come most psychotics don't seem to be manic – that is, why aren't they overconfident in anything except their delusions?)

LRF discuss prediction error in terms of mild surprise and annoyance; you didn't expect a beeping noise, the beeping noise happened, so you become startled. CFF discuss prediction error as sudden surprising salience, but then say that the attribution of salience to an odd stimulus creates a delusion of reference, a belief that it's somehow pregnant with secret messages. These are two very different views of prediction error; an autist wearing uncomfortable clothes might be constantly focusing on their itchiness rather than on whatever she's trying to do at the time, but she's not going to start thinking they're a sign from God. What's the difference?

Finally, although they highlighted a selection of drugs that make sense within their model, others seem not to. For example, there's some discussion of [ampakines for schizophrenia](#). But this is the opposite of what you'd want if psychosis involved overactive AMPA signaling! I'm not saying that the ampakines for schizophrenia definitely work, but they don't seem to make the schizophrenia noticeably worse either.

Probably this will end the same way most things in psychiatry end – hopelessly bogged down in complexity. Probably AMPA does one thing in one part of the brain, the opposite in other parts of the brain, and it's all nonlinear and different amounts of AMPA will have totally different effects and maybe downregulate itself somewhere else.

Still, it's neat to have at least a vague high-level overview of what *might* be going on.

Book Review: Surfing Uncertainty

Posted on September 5, 2017 by Scott Alexander



Related to: [It's Bayes All The Way Up](#), [Why Are Transgender People Immune To Optical Illusions?](#), [Can We Link Perception And Cognition?](#)

I

Sometimes I have the fantasy of being able to glut myself on Knowledge. I imagine meeting a time traveler from 2500, who takes pity on me and gives me a book from the future where all my questions have been answered, one after another. What's consciousness? That's in Chapter 5. How did something arise out of nothing? Chapter 7. It all makes perfect intuitive sense and is fully vouched by unimpeachable authorities. I assume something like this is how everyone spends their first couple of days in Heaven, whatever it is they do for the rest of Eternity.

And every so often, my fantasy comes true. Not by time travel or divine intervention, but by failing so badly at paying attention to the literature that by the time I realize people are working on a problem it's already been investigated, experimented upon, organized into a paradigm, tested, and then placed in a nice package and wrapped up with a pretty pink bow so I can enjoy it all at once.

The predictive processing model is one of these well-wrapped packages. Unbeknownst to me, over the past decade or so neuroscientists have come up with a real *theory* of how the brain works – a real unifying framework theory like Darwin's or Einstein's – and it's beautiful and it makes complete sense.

[Surfing Uncertainty](#) isn't pop science and isn't easy reading. Sometimes it's on the border of possible-at-all reading. Author Andy Clark (a professor of logic and metaphysics, of all things!) is clearly brilliant, but prone to going on long digressions about various esoteric philosophy-of-cognitive-science debates. In particular, he's obsessed with showing how "embodied" everything is all the time. This gets kind of awkward, since the predictive processing model isn't really a natural match for embodiment theory, and describes a brain which is pretty embodied in some ways but not-so-embodied in others. If you want a hundred pages of apologia along the lines of "this may not *look* embodied, but if you squint you'll see how super-duper embodied it really is!", this is your book.

It's also your book if you want to learn about predictive processing at all, since as far as I know this is the only existing book-length treatment of the subject. And it's comprehensive, scholarly, and very good at giving a good introduction to the theory and why it's so important. So let's be grateful for what we've got and take a look.

II

Stanislas Dehaene writes of our senses:

We never see the world as our retina sees it. In fact, it would be a pretty horrible sight: a highly distorted set of light and dark pixels, blown up toward the center of the retina, masked by blood vessels, with a massive hole at the location of the "blind spot" where cables leave for the brain; the image would constantly

blur and change as our gaze moved around. What we see, instead, is a three-dimensional scene, corrected for retinal defects, mended at the blind spot, stabilized for our eye and head movements, and massively reinterpreted based on our previous experience of similar visual scenes. All these operations unfold unconsciously—although many of them are so complicated that they resist computer modeling. For instance, our visual system detects the presence of shadows in the image and removes them. At a glance, our brain unconsciously infers the sources of lights and deduces the shape, opacity, reflectance, and luminance of the objects.

Predictive processing begins by asking: how does this happen? By what process do our incomprehensible sense-data get turned into a meaningful picture of the world?

The key insight: the brain is a multi-layer prediction machine. All neural processing consists of two streams: a bottom-up stream of sense data, and a top-down stream of predictions. These streams interface at each level of processing, comparing themselves to each other and adjusting themselves as necessary.

The bottom-up stream starts out as all that incomprehensible light and darkness and noise that we need to process. It gradually moves up all the cognitive layers that we already knew existed – the edge-detectors that resolve it into edges, the object-detectors that shape the edges into solid objects, et cetera.

The top-down stream starts with everything you know about the world, all your best heuristics, all your priors, everything that's ever happened to you before – everything from “solid objects can't pass through one another” to “ $e = mc^2$ ” to “that guy in the blue uniform is probably a policeman”. It uses its knowledge of concepts to make predictions – not in the form of verbal statements, but in the form of expected sense data. It makes some guesses about what you're going to see, hear, and feel next, and asks “Like this?” These predictions gradually move *down* all the cognitive layers to generate lower-level predictions. If that uniformed guy was a policeman, how would that affect the various objects in the scene? Given the answer to that question, how would it affect the distribution of edges in the scene? Given the answer to *that* question, how would it affect the raw-sense data received?

Both streams are probabilistic in nature. The bottom-up sensory stream has to deal with fog, static, darkness, and neural noise; it knows that whatever forms it tries to extract from this signal might or might not be real. For its part, the top-down predictive stream knows that predicting the future is inherently difficult and its models are often flawed. So both streams contain not only data but estimates of the precision of that data. A bottom-up percept of an elephant right in front of you on a clear day might be labelled “very high precision”; one of a vague form in a swirling mist far away might be labelled “very low precision”. A top-down prediction that water will be wet might be labelled “very high precision”; one that the stock market will go up might be labelled “very low precision”.

As these two streams move through the brain side-by-side, they continually interface with each other. Each level receives the predictions from the level above it and the sense data from the level below it. Then each level uses [Bayes' Theorem](#) to integrate these two sources of probabilistic evidence as best it can. This can end up a couple of different ways.

First, the sense data and predictions may more-or-less match. In this case, the layer stays quiet, indicating “all is well”, and the higher layers never even hear about it. The higher levels just keep predicting whatever they were predicting before.

Second, low-precision sense data might contradict high-precision predictions. The Bayesian math will conclude that the predictions are still probably right, but the sense data are wrong. The lower levels will “cook the books” – rewrite the sense data to make it look as predicted – and then continue to be quiet and signal that all is well. The higher levels continue to stick to their predictions.

Third, there might be some unresolvable conflict between high-precision sense-data and predictions. The Bayesian math will indicate that the predictions are probably wrong. The neurons involved will fire, indicating “surprisal” – a gratuitously-technical neuroscience term for surprise. The higher the degree of mismatch, and the higher the supposed precision of the data that led to the mismatch, the more surprisal – and the louder the alarm sent to the higher levels.

When the higher levels receive the alarms from the lower levels, *this is their equivalent of bottom-up sense-data*. They ask themselves: “Did the even-higher-levels predict this would happen?” If so, they themselves stay quiet. If not, they might try to change their own models that map higher-level predictions to lower-level sense data. Or they might try to cook the books themselves to smooth over the discrepancy. If none of this works, they send alarms to the even-higher-levels.

All the levels really hate hearing alarms. Their goal is to *minimize surprisal* – to become so good at predicting the world (conditional on the predictions sent by higher levels) that nothing ever surprises them. Surprise prompts a frenzy of activity adjusting the parameters of models – or deploying new models – until the surprise stops.

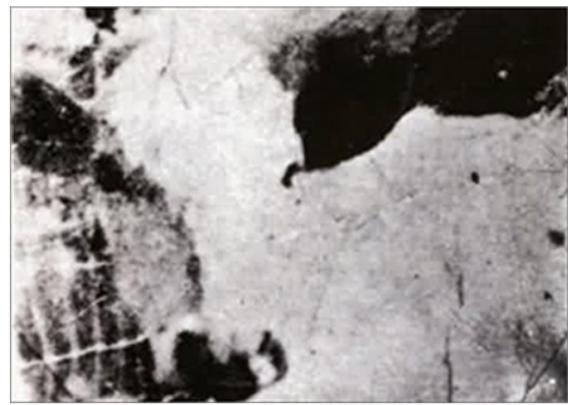
All of this happens several times a second. The lower levels constantly shoot sense data at the upper levels, which constantly adjust their hypotheses and shoot them down at the lower levels. When surprise is registered, the relevant levels change their hypotheses or pass the buck upwards. After umpteen zillion cycles, everyone has the right hypotheses, nobody is surprised by anything, and the brain rests and moves on to the next task. As per the book:

To deal rapidly and fluently with an uncertain and noisy world, brains like ours have become masters of prediction – surfing the waves and noisy and ambiguous sensory stimulation by, in effect, trying to stay just ahead of them. A skilled surfer stays ‘in the pocket’: close to, yet just ahead of the place where the wave is breaking. This provides power and, when the wave breaks, it does not catch her. The brain’s task is not dissimilar. By constantly attempting to predict the incoming sensory signal we become able – in ways we shall soon explore in detail – to learn about the world around us and to engage that world in thought and action.

The result is perception, which the PP theory describes as “controlled hallucination”. You’re not seeing the world as it is, exactly. You’re seeing your predictions about the world, cashed out as expected sensations, then shaped/constrained by the actual sense data.

III

Enough talk. Let’s give some examples. Most of you have probably seen these before, but it never hurts to remind:



This demonstrates the degree to which the brain depends on top-down hypotheses to make sense of the bottom-up data. To most people, these two pictures start off looking like incoherent blotches of light and darkness. Once they figure out what they are ([spoiler](#)) the scene becomes obvious and coherent. According to the predictive processing model, this is how we perceive everything all the time – except usually the concepts necessary to make the scene fit together come from our higher-level predictions instead of from clicking on a spoiler link.



This demonstrates how the top-down stream's efforts to shape the bottom-up stream and make it more coherent can sometimes "cook the books" and alter sensation entirely. The real picture says "PARIS IN THE THE SPRINGTIME" (note the duplicated word "the"!). The top-down stream predicts this should be a meaningful sentence that obeys English grammar, and so replaces the the bottom-up stream with what it thinks that it *should* have said. This is a very powerful process – how many times have I repeated the the word "the" in this paragraph alone without you noticing?

You can porabbly raed tihs ptetry wlel eevn
tohguh it's all jubemld up.

A more ambiguous example of “perception as controlled hallucination”. Here your experience doesn’t quite *deny* the jumbled-up nature of the letters, but it superimposes a “better” and more coherent experience which appears naturally alongside.

<https://www.youtube.com/embed/Ftdb5EKqjlo?rel=0&controls=0&showinfo=0&start=42>

Next up – this low-quality video of an airplane flying at night. Notice how after an instant, you start to predict the movement and characteristics of the airplane, so that you’re no longer surprised by the blinking light, the movement, the other blinking light, the camera shakiness, or anything like that – in fact, if the light *stopped* blinking, you would be surprised, even though naively nothing could be less surprising than a dark portion of the night sky staying dark. After a few seconds of this, the airplane continuing on its (pretty complicated) way just reads as “same old, same old”. Then when something else happens – like the camera panning out, or the airplane making a slight change in trajectory – you focus entirely on that, the blinking lights and movement entirely forgotten or at least packed up into “airplane continues on its blinky way”. Meanwhile, other things – like the feeling of your shirt against your skin – have been completely predicted away and blocked from consciousness, freeing you to concentrate entirely on any subtle changes in the airplane’s motion.

https://www.youtube.com/embed/66tQR7koR_Q

In the same vein: this is Rick Astley’s “Never Going To Give You Up” repeated again and again for ten hours (you can find some *weird* stuff on YouTube). The first hour, maybe you find yourself humming along occasionally. By the second hour, maybe it’s gotten kind of annoying. By the third hour, you’ve completely forgotten it’s even on at all.

But suppose that one time, somewhere around the sixth hour, it skipped two notes – just the two syllables “never”, so that Rick said “Gonna give you up.” Wouldn’t the silence where those two syllables should be sound as jarring as if somebody set off a bomb right beside you? Your brain, having predicted sounds consistent with “Never Gonna Give You Up” going on forever, suddenly finds its expectations violated and sends all sorts of alarms to the higher levels, where they eventually reach your consciousness and make you go “What the heck?”

IV

Okay. You’ve read a lot of words. You’ve looked at a lot of pictures. You’ve listened to “Never Gonna Give You Up” for ten hours. Time for the payoff. Let’s use this theory to explain everything.

1. Attention

In PP, attention measures “the confidence interval of your predictions”. Sense-data within the confidence intervals counts as a match and doesn’t register surprisal. Sense-data outside the confidence intervals fails and alerts higher levels and eventually consciousness.

This modulates the balance between the top-down and bottom-up streams. High attention means that perception is mostly based on the bottom-up stream, since every little deviation is registering an error and so the overall perceptual picture is highly constrained by sensation. Low attention means that perception is mostly based on the top-down stream, and you’re perceiving only a vague outline of the sensory image with your predictions filling in the rest.

There’s a famous experiment which you can try below – if you’re trying it, make sure to play the whole video before moving on:

<https://www.youtube.com/embed/vJG698U2Mvo?rel=0&controls=0&showinfo=0>

...

...

About half of subjects, told to watch the players passing the ball, don’t notice the gorilla. Their view of the ball-passing is closely constrained by the bottom-up stream; they see mostly what is there. But their view of the gorilla is mostly dependent on the top-down stream. Their confidence intervals are wide. Somewhere in your brain is a neuron saying “is that a guy in a gorilla suit?” Then it consults the top-down stream, which says “This is a basketball game, you moron”, and it smooths out the anomalous perception into something that makes sense like another basketball player.

But if you watch the video with the prompt “Look for something strange happening in the midst of all this basketball-playing”, you see the gorilla immediately. Your confidence intervals for unusual things are razor-thin; as soon as that neuron sees the gorilla it sends alarms to higher levels, and the higher levels quickly come up with a suitable hypothesis (“there’s a guy in a gorilla suit here”) which makes sense of the new data.

There’s an interesting analogy to vision here, where the center of your vision is very clear, and the outsides are filled in in a top-down way – I have a vague sense that my water bottle is in the periphery right now, but only because I kind of already know that, and it’s more of a mental note of “water bottle here as long as you ask no further questions” than a clear image of it. The extreme version of this is [the blind spot](#), which gets filled in entirely with predicted imagery despite receiving no sensation at all.

2. Imagination, Simulation, Dreaming, Etc.

Imagine a house. Now imagine a meteor crashing into the house. Your internal mental simulation was probably pretty good. Without even thinking about it, you got it to obey accurate physical laws like “the meteor continues on a constant trajectory”, “the impact happens in a realistic way”, “the impact shatters the meteorite”, and “the meteorite doesn’t bounce back up to space like a basketball”. Think how surprising this is.

In fact, think how surprising it is that you can imagine the house at all. This really high level concept – “house” – has been transformed in your visual imaginarium into a pretty good picture of a house, complete with various features, edges, colors, et cetera (if it hasn’t, read [here](#)). This is near-miraculous. Why do our brains have this apparently useless talent?

PP says that the highest levels of our brain make predictions *in the form of sense data*. They’re not just saying “I predict

that guy over there is a policeman", they're generating the image of a policeman, cashing it out in terms of sense data, and colliding it against the sensory stream to see how it fits. The sensory stream gradually modulates it to fit the bottom-up evidence – a white or black policeman, a mustached or clean-shaven policeman. But the top-down stream is doing a lot of the work here. We are able to imagine the meteor, using the same machinery that would guide our perception of the meteor if we saw it up in the sky.

All of this goes double for dreaming. If "perception is controlled hallucination" caused by the top-down drivers of perception constrained by bottom-up evidence, then dreams are those top-down drivers playing around with themselves unconstrained by anything at all (or else very weakly constrained by bottom-up evidence, like when it's really cold in your bedroom and you dream you're exploring the North Pole).

A lot of people claim higher levels of this – lucid dreaming, astral projection, you name it, worlds exactly as convincing as our own but entirely imaginary. Predictive processing is very sympathetic to these accounts. The generative models that create predictions are really good; they can simulate the world well enough that it rarely surprises us. They also connect through various layers to our bottom-level perceptual apparatus, cashing out their predictions in terms of the lowest-level sensory signals. Given that we've got a top-notch world-simulator plus perception-generator in our heads, it shouldn't be surprising when we occasionally perceive ourselves in simulated worlds.

3. Priming

I don't mean the weird made-up kinds of priming that don't replicate. I mean the very firmly established ones, like the one where, if you flash the word "DOCTOR" at a subject, they'll be much faster and more skillful in decoding a series of jumbled and blurred letters into the word "NURSE".

This is classic predictive processing. The top-down stream's whole job is to assist the bottom-up stream in making sense of complicated fuzzy sensory data. After it hears the word "DOCTOR", the top-down stream is already thinking "Okay, so we're talking about health care professionals". This creeps through all the lower levels as a prior for health-care related things; when the sense organs receive data that can be associated in a health-care related manner, the high prior helps increase the precision of this possibility until it immediately becomes the overwhelming leading hypothesis.

4. Learning

There's a philosophical debate – which I'm not too familiar with, so sorry if I get it wrong – about how "unsupervised learning" is possible. Supervised reinforcement learning is when an agent tries various stuff, and then someone tells the agent if it's right or wrong. Unsupervised learning is when nobody's around to tell you, and it's what humans do all the time.

PP offers a compelling explanation: we create models that generate sense data, and keep those models if the generated sense data match observation. Models that predict sense data well stick around; models that fail to predict the sense data accurately get thrown out. Because of all those lower layers adjusting out contingent features of the sensory stream, any given model is left with exactly the sense data necessary to tell it whether it's right or wrong.

PP isn't *exactly* blank slate, but it's compatible with a slate that's pretty fricking blank. Clark discusses "hyperpriors" – extremely basic assumptions about the world that we probably need to make sense of anything at all. For example, one hyperprior is sensory synchronicity – the idea that our five different senses are describing the same world, and that the stereo we see might be the source of the music we hear. Another hyperprior is object permanence – the idea that the world is divided into specific objects that stick around whether or not they're in the sensory field. Clark says that some hyperpriors *might* be innate – but says they don't have to be, since PP is strong enough to learn them on its

own if it has to. For example, after enough examples of, say, seeing a stereo being smashed with a hammer at the same time that music suddenly stops, the brain can infer that connecting the visual and auditory evidence together is a useful hack that helps it to predict the sensory stream.

I can't help thinking here of [Molyneux's Problem](#), a thought experiment about a blind-from-birth person who navigates the world through touch alone. If suddenly given sight, could the blind person naturally connect the visual appearance of a cube to her own concept "cube", which she derived from the way cubes feel? In 2003, some researchers took advantage of a new cutting-edge blindness treatment to [test this out](#); they found that no, the link isn't intuitively obvious to them. Score one for learned hyperpriors.

But learning goes all the way from these kinds of really basic hyperpriors all the way up to normal learning like what the capital of France is – which, if nothing else, helps predict what's going to be on the other side of your geography flashcard, and which high-level systems might keep as a useful concept to help it make sense of the world and predict events.

5. Motor Behavior

About a third of *Surfing Uncertainty* is on the motor system, it mostly didn't seem that interesting to me, and I don't have time to do it justice here (I might make another post on one especially interesting point). But this has been kind of ignored so far. If the brain is mostly just in the business of making predictions, what exactly is the motor system doing?

Based on a bunch of really excellent experiments that I don't have time to describe here, Clark concludes: it's predicting action, which causes the action to happen.

This part is almost funny. Remember, the brain really hates prediction error and does its best to minimize it. With failed predictions about eg vision, there's not much you can do except change your models and try to predict better next time. But with predictions about proprioceptive sense data (ie your sense of where your joints are), there's an easy way to resolve prediction error: just move your joints so they match the prediction. So (and I'm asserting this, but see Chapters 4 and 5 of the book to hear the scientific case for this position) if you want to lift your arm, your brain just predicts *really really strongly* that your arm has been lifted, and then lets the lower levels' drive to minimize prediction error do the rest.

Under this model, the "prediction" of a movement isn't just the idle thought that a movement might occur, it's *the actual motor program*. This gets unpacked at all the various layers – joint sense, proprioception, the exact tension level of various muscles – and finally ends up in a particular fluid movement:

Friston and colleagues... suggest that precise proprioceptive predictions directly elicit motor actions. This means that motor commands have been replaced by (or as I would rather say, implemented by) proprioceptive predictions. According to active inference, the agent moves body and sensors in ways that amount to actively seeking out the sensory consequences that their brains expect. Perception, cognition, and action – if this unifying perspective proves correct – work together to minimize sensory prediction errors by selectively sampling and actively sculpting the stimulus array. This erases any fundamental computational line between perception and the control of action. There remains [only] an obvious difference in direction of fit. Perception here matches hural hypotheses to sensory inputs... while action brings unfolding proprioceptive inputs into line with neural predictions. The difference, as Anscombe famously remarked, is akin to that between consulting a shopping list (thus letting the list determine the contents of the shopping basket) and listing some actually purchased items (thus letting the contents of the shopping basket determine the list). But despite the difference in direction of fit, the underlying form of the neural computations is now revealed as the same.

6. Tickling Yourself

One consequence of the PP model is that organisms are continually adjusting out their own actions. For example, if you're trying to predict the movement of an antelope you're chasing across the visual field, you need to adjust out the up-down motion of your own running. So one "hyperprior" that the body probably learns pretty early is that if it itself makes a motion, it should expect to feel the consequences of that motion.

There's a really interesting illusion called the force-matching task. A researcher exerts some force against a subject, then asks the subject to exert exactly that much force against something else. Subjects' forces are usually biased upwards – they exert more force than they were supposed to – probably because their brain's prediction engines are "cancelling out" their own force. Clark describes one interesting implication:

The same pair of mechanisms (forward-model-based prediction and the dampening of resulting well-predicted sensation) have been invoked to explain the unsettling phenomenon of 'force escalation'. In force escalation, physical exchanges (playground fights being the most common exemplar) mutually ramp up via a kind of step-ladder effect in which each person believes the other one hit them harder. Shergill et al describe experiments that suggest that in such cases each person is truthfully reporting their own sensations, but that those sensations are skewed by the attenuating effects of self-prediction. Thus, 'self-generated forces are perceived as weaker than externally generated forces of equal magnitude.'

This also explains why you can't tickle yourself – your body predicts and adjusts away your own actions, leaving only an attenuated version.

7. The Placebo Effect

We hear a lot about "pain gating" in the spine, but the PP model does a good job of explaining what this is: adjusting pain based on top-down priors. If you believe you should be in pain, the brain will use that as a filter to interpret ambiguous low-precision pain signals. If you believe you shouldn't, the brain will be more likely to assume ambiguous low-precision pain signals are a mistake. So if you take a pill that doctors assure you will cure your pain, then your lower layers are more likely to interpret pain signals as noise, "cook the books" and prevent them from reaching your consciousness.

Psychosomatic pain is the opposite of this; see Section 7.10 of the book for a fuller explanation.

8. Asch Conformity Experiment

More speculative, and not from the book. But remember this one? A psychologist asked subjects which lines were the same length as other lines. The lines were all *kind of* similar lengths, but most subjects were still able to get the right answer. Then he put the subjects in a group with confederates; all of the confederates gave the same wrong answer. When the subject's turn came, usually they would disbelieve their eyes and give the same wrong answer as the confederates.

The bottom-up stream provided some ambiguous low-precision bottom-up evidence pointing toward one line. But in the final Bayesian computation, those were swamped by the strong top-down prediction that it would be another. So the middle layers "cooked the books" and replaced the perceived sensation with the predicted one. From Wikipedia:

Participants who conformed to the majority on at least 50% of trials reported reacting with what Asch called a "distortion of perception". These participants, who made up a distinct minority (only 12 subjects),

expressed the belief that the confederates' answers were correct, and were apparently unaware that the majority were giving incorrect answers.

9. Neurochemistry

PP offers a way to a psychopharmacological holy grail – an explanation of what different neurotransmitters really *mean*, on a human-comprehensible level. Previous attempts to do this, like “dopamine represents reward, serotonin represents calmness”, have been so wildly inadequate that the whole question seems kind of disreputable these days.

But as per PP, the NMDA glutamatergic system mostly carries the top-down stream, the AMPA glutamatergic system mostly carries the bottom-up stream, and dopamine mostly carries something related to precision, confidence intervals, and surprisal levels. This matches a lot of observational data in a weirdly consistent way – for example, it doesn't take a lot of imagination to think of the slow, hesitant movements of Parkinson's disease as having “low motor confidence”.

10. Autism

Various research in the PP tradition has coalesced around the idea of autism as an unusually high reliance on bottom-up rather than top-down information, leading to “weak central coherence” and constant surprisal as the sensory data fails to fall within pathologically narrow confidence intervals.

Autistic people classically [can't stand tags on clothing](#) – they find them too scratchy and annoying. Remember the example from Part III about how you successfully predicted away the feeling of the shirt on your back, and so manage never to think about it when you're trying to concentrate on more important things? Autistic people can't do that as well. Even though they have a layer in their brain predicting “will continue to feel shirt”, the prediction is too precise; it predicts that next second, the shirt will produce *exactly* the same pattern of sensations it does now. But realistically as you move around or catch passing breezes the shirt will change ever so slightly – at which point autistic people's brains will send alarms all the way up to consciousness, and they'll perceive it as “my shirt is annoying”.

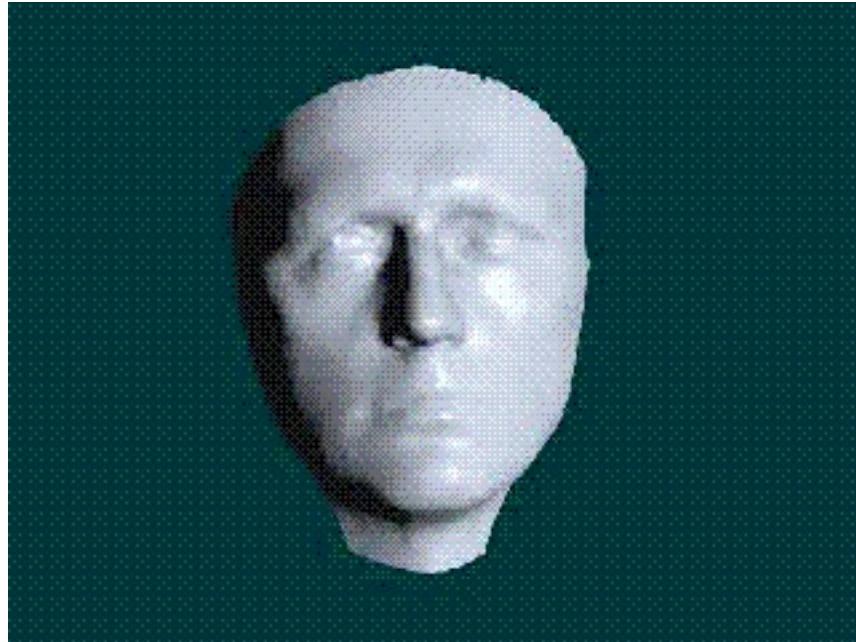
Or consider the classic autistic demand for routine, and misery as soon as the routine is disrupted. Because their brains can only make very precise predictions, the slightest disruption to routine registers as strong surprisal, strong prediction failure, and “oh no, all of my models have failed, nothing is true, anything is possible!” Compare to a neurotypical person in the same situation, who would just relax their confidence intervals a little bit and say “Okay, this is basically 99% like a normal day, whatever”. It would take something genuinely unpredictable – like being thrown on an unexplored continent or something – to give these people the same feeling of surprise and unpredictability.

This model also predicts autistic people's strengths. We know that polygenic risk for autism is [positively associated with IQ](#). This would make sense if the central feature of autism was a sort of increased mental precision. It would also help explain why autistic people seem to excel in high-need-for-precision areas like mathematics and computer programming.

11. Schizophrenia

Converging lines of research suggest this also involves weak priors, apparently at a different level to autism and with different results after various compensatory mechanisms have had their chance to kick in. One especially interesting study asked neurotypicals and schizophrenics to follow a moving light, much like the airplane video in Part III above. When the light moved in a predictable pattern, the neurotypicals were much better at tracking it; when it was a deliberately perverse video specifically designed to frustrate expectations, the schizophrenics actually did better. This

suggests that neurotypicals were guided by correct top-down priors about where the light would be going; schizophrenics had very weak priors and so weren't really guided very well, but also didn't screw up when the light did something unpredictable. Schizophrenics are also famous for not being fooled by the "hollow mask" (below) and other illusions where top-down predictions falsely constrain bottom-up evidence. My guess is they'd be more likely to see both 'the's in the "PARIS IN THE SPRINGTIME" image above.



The exact route from this sort of thing to schizophrenia is really complicated, and anyone interested should check out Section 2.12 and the whole of Chapter 7 from the book. But the basic story is that it creates waves of anomalous prediction error and surprisal, leading to the so-called "delusions of significance" where schizophrenics believe that eg the fact that someone is wearing a hat is some sort of incredibly important cosmic message. Schizophrenics' brains try to produce hypotheses that explain all of these prediction errors and reduce surprise – which is impossible, because the prediction errors are random. This results in incredibly weird hypotheses, and eventually in schizophrenic brains being willing to ignore the bottom-up stream entirely – hence hallucinations.

All this is treated with antipsychotics, which antagonize dopamine, which – remember – represents confidence level. So basically the medication is telling the brain "YOU CAN IGNORE ALL THIS PREDICTION ERROR, EVERYTHING YOU'RE PERCEIVING IS TOTALLY GARBAGE SPURIOUS DATA" – which turns out to be exactly the message it needs to hear.

An interesting corollary of all this – because all of schizophrenics' predictive models are so screwy, they lose the ability to use the "adjust away the consequences of your own actions" hack discussed in Part 5 of this section. That means their own actions *don't* get predicted out, and seem like the actions of a foreign agent. This is why they get so-called "delusions of agency", like "the government beamed that thought into my brain" or "aliens caused my arm to move just now". And in case you were wondering – [yes, schizophrenics can tickle themselves](#).

12. Everything else

I can't possibly do justice to the whole of *Surfing Uncertainty*, which includes sections in which it provides lucid and compelling PP-based explanations of hallucinations, binocular rivalry, conflict escalation, and various optical illusions. More speculatively, I can think of really interesting connections to things like phantom limbs, creativity (and its

association with certain mental disorders), depression, meditation, etc, etc, etc.

The general rule in psychiatry is: if you think you've found a theory that explains everything, diagnose yourself with mania and check yourself into the hospital. Maybe I'm not at that point yet – for example, I don't think PP does anything to explain what mania itself is. But I'm pretty close.

IV

This is a really poor book review of *Surfing Uncertainty*, because I only partly understood it. I'm leaving out a *lot* of stuff about the motor system, debate over philosophical concepts with names like "enactivism", descriptions of how neurons form and uniform coalitions, and of course a hundred pages of apologia along the lines of "this may not *look* embodied, but if you squint you'll see how super-duper embodied it really is!". As I reread and hopefully come to understand some of this better, it might show up in future posts.

But speaking of philosophical debates, there's one thing that really struck me about the PP model. [Voodoo psychology](#) suggests that culture and expectation tyrannically shape our perceptions. Taken to an extreme, objective knowledge is impossible, since all our sense-data is filtered through our own bias. Taken to a *very far* extreme, we get things like [What The !@#\\$ Do We Know?](#)'s claim that the Native Americans literally couldn't see Columbus' ships, because they had no concept of "caravel" and so the percept just failed to register. This sort of thing tends to end by arguing that science was invented by straight white men, and so probably just reflects straight white maleness, and so we should ignore it completely and go frolic in the forest or something.

Predictive processing is sympathetic to all this. It takes all of this stuff like priming and the placebo effect, and it predicts it handily. But it doesn't give up. It (theoretically) puts it all on a sound mathematical footing, explaining exactly how *much* our expectations should shape our reality, and in which *ways* our expectation should shape our reality. I feel like someone armed with predictive processing and a bit of luck should have been able to predict that placebo effect and basic priming would work, but stereotype threat and social priming wouldn't. Maybe this is total retrodictive cheating. But I feel like it should be possible.

If this is true, it gives us more confidence that our perceptions should correspond – at least a little – to the external world. We can accept that we may be misreading “PARIS IN THE SPRINGTIME” while remaining confident that we *wouldn’t* misread “PARIS IN THE SPRINGTIME” as containing only one “the”. Top-down processing very occasionally meddles in bottom-up sensation, but (as long as you’re not schizophrenic), it sticks to an advisory role rather than being able to steamroll over arbitrary amounts of reality.

The rationalist project is overcoming bias, and that requires both an admission that bias is possible, and a hope that there's something *other* than bias which we can latch onto as a guide. Predictive processing gives us more confidence in both, and helps provide a convincing framework we can use to figure out what's going on at all levels of cognition.

Book Review: Behavior – The Control Of Perception

Posted on March 6, 2017 by Scott Alexander



Epistemic status: I only partly understood this book and am trying to review it anyway as best I can.

I

People complain that psychology is paradigmless; it never got its Darwin or Newton to tie everything together. Nowadays people are pretty relaxed about that; who needs paradigms when you can do n = 50 studies on a mildly interesting effect? But historically, there were all of these larger-than-life figures who were sure they'd found the paradigm, geniuses who founded schools which flourished for a while, made big promises, then either fizzled out or toned down their claims enough to be accepted as slightly kooky parts of the mainstream. Sigmund Freud. BF Skinner. Carl Rogers. And those are just the big ones close to the mainstream. Everyone from Ayn Rand to Scientology tried their hand at the paradigm-inventing business for a while.

Will Powers (whose name turns out to be pretty appropriate) lands somewhere in the middle of this pack. He was an engineer/inventor who specialized in cybernetic systems but wandered into psychology sometime in the sixties. He argued that everything in the brain made perfect sense if you understood cybernetic principles, and came up with a very complicated but all-encompassing idea called Perceptual Control Theory which explained thought, sensation and behavior. A few people paid attention, and his work was described as paradigm-shifting by no less of an expert on paradigm shifts than Thomas Kuhn. But in the end it never really went anywhere, psychology moved on, and nowadays only a handful of people continue research in his tradition.

Somehow I kept running into this handful, and they kept telling me to read Powers' book [Behavior: The Control Of Perception](#), and I keep avoiding it. A few weeks ago I was driving down the road and I had a moment of introspection where I realized everything I was doing exactly fit Powers' theory, so I decided to give it a chance.

Powers specializes in *control systems*. The classic control system is a thermostat, which controls temperature. It has a reference point, let's say 70 degrees. If it gets much below 70 degrees, it turns on the heater until it's 70 again; if it gets much above 70 degrees, it turns on the air conditioner until it's 70 again. This is more complicated than it sounds, and there are other control systems that are even more complicated, but that's the principle. Perceptual Control Theory says that this kind of system is the basic unit of the human brain.

While I was driving on the highway a few weeks ago, I realized how much of what I do *is* perceptual control. For example, I was effortlessly maintaining the right distance from the car in front of me. If the car sped up a tiny bit, I would speed up a tiny bit. If the car slowed down a little bit, I would slow down a little bit. Likewise, I was maintaining the right angle relative to the road: if I found myself veering right, I would turn slightly to the left; if I found myself veering left, I would turn slightly to the right.

The theory goes further: while I'm in the car, I'm also operating as my own thermostat. I have a desired temperature: if I go below it, I'll turn on the heat, and if I go above it, I'll turn on the AC. I have a desired level of satiety: if I'm hungry, I'll stop and get something to eat; if I'm too full, there's maybe not a *huge* amount I can do but I'll at least stop eating. I have a desired level of light: if it's too dark, I'll turn on the lights; if it's too bright I'll put down the sun visor. I even have a desired angle to be sitting at: if I'm too far forward, I'll relax and lean back a little bit; if I'm too far back, I'll

move forwards. All of this is so easy and automatic that I never think about it.

Powers' theories go further. He agrees that my brain sets up a control system to keep my car the proper distance from the car in front of it. But how do I determine "the proper distance"? That quantity must be fed to the system by other parts of my brain. For example, suppose that the roads are icy and I know my brakes don't work very well in the ice; I might keep a much further distance than usual. I'll still be controlling the distance, I'll just be controlling it *differently*. If the brain is control systems all the way down, we can imagine a higher-tier system controlling "accident risk" at some level (presumably low, or zero) feeding a distance level into a lower-tier system controlling car distance at whatever level it receives. We can even imagine higher systems than this. Suppose I'm depressed, I've become suicidal, I want to die in a car accident, but in order not to scandalize my family I have to let the accident happen sort of naturally. I have a top-level system controlling "desire to die" which tells a middle-level system controlling "accident risk" what level it should go at (high), which in turn tells a lower-tier system controlling "car distance" what level *it* should go at (very close).

It doesn't even end there. My system controlling "car distance" is sending signals to a lower-tier system controlling muscle tension on my foot on the accelerator, giving it a new reference level (contracted muscles that push down on the accelerator really hard). Except this is an oversimplification, because everything that has to do with muscles is a million times more complicated than any reasonable person would think (at least until they play [qwop](#)) and so there's actually a big hierarchy of control systems just going from "want to go faster" to "successfully tense accelerator-related muscles".

II

Actually, Powers is at his most convincing when he talks about these lower-level functions. At this point I think it's [pretty mainstream](#) to say that muscle tension is set by a control system, with the Golgi tendon organs giving feedback and the spinal cord doing the calculations. Powers goes further (and I don't know how mainstream this next part is, but I'm guessing at least somewhat), saying that this is a first-tier control system, which is itself controlled by a second-tier "direction" control system centered in the nuclei of the brainstem, which is itself controlled by a third-tier "position" control system centered in the cerebellum/thalamus/midbrain (a friendly amendment might add the basal ganglia, which Powers doesn't seem to know much about).

If you stimulate certain parts of a cat's midbrain, it will go into specific positions – for example, a position like it's ready to pounce. So it seems like those areas "code for" position. But in order to have a neuron/area/whatever that codes for position, it needs to have hierarchical control over lots of lower-level things. For example, it needs to make sure the leg muscles are however tense they're supposed to be in a pouncing position. So the third-tier position control system controls the second-tier direction control system at whatever level is necessary to make the second-tier direction control system control the first-tier muscle control system at whatever level is necessary to get the muscles in the right position.

The fourth- and fifth-tier systems, now well into the cortex (and maybe basal ganglia again) deal with sequences, eg "walking" or "playing a certain tune on the piano". Once again, activating a fourth/fifth-tier system will activate this higher-level concept ("walking"), which alters the reference levels for a third-tier system ("getting into a certain position"), which alters a second-tier system ("moving in a certain direction"), which alters a first-tier system ("tensing/relaxing muscles").

Why do I like this theory so much? First, it correctly notes that (almost) the only thing the brain can actually *do* is change muscle tension. Yet we never think in terms of muscle tension. We don't think "I am going to tense my thigh muscle, now untense it, now tense my ankle muscle, now...", we just think "I'm going to walk". Heck, half the time we don't even think *that*, we think "I'm just going to go to the fridge" and the walking happens automatically. On the other

hand, if we really want, we *can* consciously change our position, the level of tension in a certain muscle, etc. It's just that usually we deal in higher-level abstractions that automatically carry all the lower ones along with them.

Second, it explains the structure of the brain in a way I haven't seen other things do. I always hear neuroscientists talk about "this nucleus relays signals to that nucleus" or "this structure is a way station for this other structure". Spend too much time reading that kind of stuff, and you start to think of the brain as a giant relay race, where the medulla passes signals onto the thalamus which passes it to the basal ganglia which passes it to the frontal lobe and then, suddenly, thought! The obvious question there is "why do you have so many structures that just relay things to other structures?" Sometimes neuroscientists will say "Well, some processing gets done here", or even better "Well, this system modulates that system", but they're always very vague on what exactly that means. Powers' hierarchy of fifth-tier systems passing their calculations on to fourth-tier systems and so on is exactly the sort of thing that would make sense of all this relaying. My guess is every theory of neuroscience has something at least this smart, but I'd never heard it explained this well before.

Third, it's the clearest explanation of tremors I've ever heard. Consider the thermostat above. When the temperature gets below 65, it turns on the heat until the temperature gets above 70, then stops, then waits as the hot air leaks out through the window or whatever and it's 65 again, then turns on the heat again. If we chart temperature in a room with a thermostat, it will look sort of like a sine wave or zigzag with regular up/down motions. This is a basic principle of anything being controlled by a less-than-perfect control system. Our body has microtremors all the time, but when we get brain damage or some other problem, a very common symptom is noticeable tremors. These come in many different varieties that give clues to the level of brain damage and which doctors are just told to memorize. Powers actually explains them:

When first-order systems become unstable, as when muscles exert too much effort), clonus oscillations are seen, at roughly ten cycles per second. Second-order instability, as in the tremors of Parkinsonism, involves groups of muscles and is of lower frequency, around three cycles per second or so. Third-order instability is slower still, slow enough that it can be characterized as "purpose tremor" or "over-correction". Certain cerebellar damage due to injury or disease can result in over- and under-shooting the mark during actions such as reaching out to grasp something, either in a continuous self-sustained oscillation or a slowly decreasing series of alternating movements.

This isn't perfect – for example, Parkinsonian tremor is usually caused by damage to the basal ganglia and the cortex, which is really hard to square with Powers' claim that it's caused by damage to second-tier systems in the medulla. But after reading this, it's really hard not to think of tremors as failures in control systems, or of the different types of tremor as failures in different levels of control system. For example, athetoid tremors are weird, seemingly purposeful, constant twisting movements caused by problems in the thalamus or some related system; after reading Powers, it's impossible for me not to think of them as failures in third-order control systems. This becomes especially clear if we compare to Powers' constant foil/nemesis, the Behaviorists. Stick to a stimulus-response paradigm, and there's no reason damaged brains should make weird twisting movements all the time. On a control-systems paradigm, it's *obvious* that that would happen.

There are occasional claims that perceptual control theory can predict certain things about muscles and coordination better than other theories, sometimes with absurdly high accuracy of like $r = 0.9$ or something. Powers makes some of these claims in the book, but I can't check them because I don't have the original data he worked with and I don't know how to calculate cybernetic control system outputs. But the last time I saw someone bring up one of these supposed experiments it was thoroughly shot down by people who knew more statistics. And I found a blog post where somebody who knows a lot about intricacies of muscle movement says PCT can predict some things but not much better than competing theories. In terms of predicting very specific things about human muscular movement its record seems to be kind of so-so.

III

And I start to get very skeptical when Powers moves to higher-tier control systems. His sixth tier is “relationships”, seventh is “programs”, eighth is “principles”, and ninth is “systems”. Although these tiers receive just as many pages as the earlier ones, they start sounding very abstract and they correlate a lot less well with anatomy. I understand the urge to postulate them – if you’ve already decided that the fundamental unit of the brain is the control system, why not try to explain things with control systems all the way up? – but it becomes kind of a stretch. It’s easy to see what it means to control the distance between me and the car in front of me; it’s harder to see what it means to control for “communism” or “honesty” or things like that.

I *think* the way things are supposed to work is like this. A ninth-tier system controls a very abstract concept like “communism”. So suppose you are a communist; that means your internal communism-thermostat is set to maintain your communism at a high level. That propagates down to eighth-tier principles, which are slightly less abstract concepts like “greed”; maybe your ninth-tier communism-thermostat sets your eighth-tier greed thermostat to a very low temperature because communists aren’t supposed to be greedy. Your eighth-tier greed thermostat affects levels of seventh-tier logical programs like “going to work and earning money” and “giving to charity”. I’m not really sure how the sixth-tier fits into this example, but let’s suppose that your work is hammering things. Then the fifth-tier system moves your muscles in the right sequence to hammer things, and so on with all the lower tiers as above.

Sometimes these control systems come into contact with each other. For example, suppose that along with my ninth-tier system controlling “communism”, I also have a ninth-tier system controlling “family values”; I am both an avowed communist and a family man. My family values system thinks that it’s important that I earn enough to provide for my family, so while my communism-system is trying to input a low reference level for my greed-thermostat, my family-values-system is trying to input a high one. Powers gets into some really interesting examples of what happens in real industrial cybernetic systems when two opposing high-level control systems get in a fight, and thinks this is the source of all human neurosis and akrasia. I think he later wrote a self-help book based around this (hence the nominative determinism). I am not very convinced.

Am I strawmanning this picture? I’m not sure. I think one testable consequence of it is supposed to be that if we’re really controlling for communism, in the cybernetic control system sense, then we should be able to test for that. For example, hide Lenin’s pen and paper so that he can’t write communist pamphlets, and he should start doing some other communist thing more in order to make up for it and keep his level of communism constant. I think some perceptual control theory people believe this is literally true, and propose experimental tests (or at least thought experiment tests) of perceptual control theory along these lines. This seems sketchy to me, on the grounds that if Lenin didn’t start doing other stuff, we could just say that communism wasn’t truly what he was controlling.

That is, suppose I notice Lenin eating lots of chocolate every day. I theorize that he’s controlling for chocolate, and so if I disturb the control system by eg shutting down his local chocolate store, he’ll find a way to restore equilibrium, eg by walking further to a different store. But actually, when I shut down his local chocolate store, he just eats less chocolate. In reality, he was controlling his food intake (as we all do; that’s what an obesity set point is) and when he lost access to chocolate, maybe he ate cupcakes instead and did fine.

In the same way, maybe we only think Lenin is controlling for communism, but he’s actually controlling for social status, and being a communist revolutionary is a good way to gain social status. So if we make it too hard for him to be a communist revolutionary, eg by taking away his pen and paper, maybe he’ll become a rock star instead and end up with the same level of social status.

This sort of thing seems so universal that as far as I can tell it makes these ideas of higher-tier control systems unproveable and unfalsifiable.

If there's any point to them at all, I think it's the way they express the same interesting phenomenological truth as the muscle movement tiers: we switch effortlessly between concentrating on low-level concepts and high-level concepts that make the low-level ones automatic. For example, I think "driving" is a good example of Powers' seventh tier, "programs" – it involves a predictable flowchart-like set of actions to achieve a simple goal. "The distance between me and the car in front of me" is a sixth-tier system, a "relationship". When I'm driving (focusing on my seventh-tier system), I don't consciously think at all about maintaining the right distance with the car in front of me. It just happens. This is really interesting in a philosophy of consciousness sense, and Powers actually gets into qualia a bit and says some things that seem a lot wiser and more moving-part-ful than most people on the subject.

It does seem like there's something going on where my decision to drive activates a lot of carefully-trained subsystems that handle the rest of it automatically, and that there's probably some neural correlate to it. But I don't know whether control systems are the right way to think about this, and I definitely don't know whether there's a sense in which "communism" is a control system.

IV

There are also some sections about things like learning and memory, which looks suspiciously like flowcharts of control systems with boxes marked "LEARNING" and "MEMORY" in them.

But I realized halfway through that I was being too harsh. Perceptual control theory wasn't quite a proposal for a new paradigm out of nowhere. It was a reaction to Behaviorism, which was still the dominant paradigm when Powers was writing. His "everything is a control system" is an attempt to improve on "everything is stimulus-response", and it really does.

For example, his theory of learning involves reward and punishment, where reward is reducing the error in a control system and punishment is increasing it. That is, suppose that you're controlling temperature, and it's too hot out. A refreshing cool glass of water would be an effective reward (since it brings you closer to your temperature reference level), and setting your hand on fire would be an effective punishment (since it brings you further from your temperature reference level). Powers notes that this explains many things Behaviorism can't. For example, they like to talk about how sugar water is a reward. But eventually rats get tired of sugar water and stop drinking it. So it seems that sugar water isn't a reward per se; it's more like reducing error in your how-much-sugar-water-should-I-have-and-did-I-already-have-the-right-amount system is the reward. If your optimal level of sugar water per day is 10 ml, then anything up to 10 ml will be a reward, and after that it will stop being attractive / start being a punishment.

As a "theory of learning", this is sort of crappy, in that I was expecting stuff about Hebb and connectionism and how memories are stored in the brain. But if you're living in an era where everybody thinks "The response to a stimulus is predictable through patterns of reward and punishment" is an A+++ Nobel-Prize-worthy learning theory, then perceptual control-based theories of learning start sounding pretty good.

So I guess it's important to see this as a product of its times. And I don't understand those times – why Behaviorism ever seemed attractive is a mystery to me, maybe requiring more backwards-reading than I can manage right now.

How useful is this book? I guess that depends on how metaphorical you want to be. Is the brain a control system? I don't know. Are police a control system trying to control crime? Are police a "response" to the "stimulus" of crime? Is a stimulus-response pairing a control system controlling for the quantity of always making sure the stimulus has the response? I think it's interesting and helpful to think of some psychological functions with these metaphors. But I'm not sure where to go from there. I think maybe there are some obvious parallels, maybe even parallels that bear fruit in empirical results, in lower level systems like motor control. Once you get to high-level systems like communism or social desirability, I'm not sure we're doing much better than the police-as-control-system metaphor. Still, I think that

it's potentially a useful concept to have.

Predictive Processing And Perceptual Control

Posted on September 6, 2017 by Scott Alexander



Yesterday's review of *Surfing Uncertainty* mentioned how predictive processing attributes movement to strong predictions about proprioceptive sensations. Because the brain tries to minimize predictive error, it moves the limbs into the positions needed to produce those sensations, fulfilling its own prophecy.

This was a really difficult concept for me to understand at first. But there were a couple of passages that helped me make an important connection. See if you start thinking the same thing I'm thinking:

To make [bodily] action come about, the motor plant behaves (Friston, Daunizeau, et al, 2010) in ways that cancel out proprioceptive prediction errors. This works because the proprioceptive prediction errors signal the difference between how the bodily plant is currently disposed and how it would be disposed were the desired actions being performed. Proprioceptive prediction error will yield (moment-by-moment) the projected proprioceptive inputs. In this way, predictions of the unfolding proprioceptive patterns that would be associated with the performance of some action actually bring that action about. This kind of scenario is neatly captured by Hawkins and Blakeslee (2004), who write that: "As strange as it sounds, when your own behavior is involved, your predictions not only precede sensation, they determine sensation."

And:

PP thus implements the distinctive circular dynamics described by Cisek and Kalaska using a famous quote from the American pragmatist John Dewey. Dewey rejects the 'passive' model of stimuli evoking responses in favour of an active and circular model in which 'the motor response determines the stimulus, just as truly as sensory stimulus determines movement'

Still not getting it? What about:

According to active inference, the agent moves body and sensors in ways that amount to actively seeking out the sensory consequences that their brains expect.

This is the model from Will Powers' [Behavior: The Control Of Perception](#).

Clark knows this. A few pages after all these quotes, he writes:

One signature of this kind of grip-based non-reconstructive dance is that it suggests a potent reversal of our ordinary way of thinking about the relations between perception and action. Instead of seeing perception as the control of action, it becomes fruitful to think of action as the control of perception [Powers 1973, Powers et al, 2011].

But I feel like this connection should be given more weight. Powers' perceptual control theory presages predictive

processing theory in a lot of ways. In particular, both share the idea of cognitive “layers”, which act at various levels (light-intensity-detection vs. edge-detection vs. object-detection, or movements vs. positions-in-space vs. specific-muscle-actions vs. specific-muscle-fiber-tensions). Upper layers decide what stimuli they want lower levels to be perceiving, and lower layers arrange themselves in the way that produce those stimuli. PCT talks about “set points” for cybernetic systems, and PP talks about “predictions”, but they both seem to be groping at the same thing.

I was least convinced by the part of PCT which represented the uppermost layers of the brain as control systems controlling various quantities like “love” or “communism”, and which sometimes seemed to veer into self-parody. PP offers an alternative by describing those layers as making predictions (sometimes “active predictions” of the sort that guide behavior) and trying to minimize predictive error. This allows lower level systems to “control for” deviation from a specific plan, rather than just monitoring the amount of some scalar quantity.

My review of *Behavior: The Control Of Perception* ended by saying:

It does seem like there's something going on where my decision to drive activates a lot of carefully-trained subsystems that handle the rest of it automatically, and that there's probably some neural correlate to it. But I don't know whether control systems are the right way to think about this... I think maybe there are some obvious parallels, maybe even parallels that bear fruit in empirical results, in lower level systems like motor control. Once you get to high-level systems like communism or social desirability, I'm not sure we're doing much better than [strained control-related metaphors].

I think my instincts were right. PCT is a good model, but what's good about it is that it approximates PP. It approximates PP best at the lower levels, and so is most useful there; its thoughts on the higher levels remain useful but start to diverge and so become less profound.

The Greek atomists like Epicurus have been totally superseded by modern atomic theory, but they still get a sort of “how did they do that?” award for using vague intuition and good instincts to cook up a scientific theory that couldn’t be proven or universally accepted until centuries later. If PP proves right, then Will Powers and PCT deserve a place in the pantheon besides them. There’s something kind of wasteful about this – we can’t properly acknowledge the cutting-edgeness of their contribution until it’s obsolete – but at the very least we can look through their other work and see if they’ve got even *more* smart ideas that might be ahead of their time.

(Along with his atomic theory, Epicurus gathered a bunch of philosophers and [mathematicians](#) into a small [cult](#) around him, who lived together in [co-ed group houses](#) preaching atheism and materialism and – as per the rumors – having [orgies](#). If we’d just agreed he was [right about everything](#) from the start, we wouldn’t have had to laboriously reinvent his whole system.)

Medicine



“*For truly on countless occasions throughout my life I have had this experience; persons for a time talk pleasantly with me because of my work among the sick, in which they think me very well trained, but when they learn later on that I am also trained in mathematics, they avoid me for the most part and are no longer at all glad to be with me.*”

”

—Galen

**

[Who By Very Slow Decay](#)

[Book Review: House of God](#)

[Side Effects May Include Anything](#)

[Cost Disease In Medicine: The Practical Perspective](#)

[The Amish Health Care System](#)

Who By Very Slow Decay

Posted on July 17, 2013 by Scott Alexander



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 “mmmmrrrgggg graaaaaaaaaaaHAAACK!”

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-mouthing doctor gloats,
and notes each new and nerve-twisted 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-ridden 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.

Book Review: House of God

Posted on November 10, 2016 by Scott Alexander



I

I'm not a big fan of war movies. I liked the first few I watched. It was all downhill from there. They all seem so similar. The Part Where You Bond With Your Squadmates. The Part Where Your Gruff Sergeant Turns Out To Have A Heart After All. The Part Where Your Friend Dies But You Have To Keep Going Anyway. The Part That Consists Of A Stirring Speech.

The problem is that war is very different from everything else, but very much like itself.

Medical internship is also very different from everything else but very much like itself. I already had two examples of it: *Scrubs* and my own experience as a medical intern (I preferred *Scrubs*). So when *every single person in the medical field* told me to read Samuel Shem's [*House of God*](#), I deferred. I deferred throughout my own internship, I deferred for another two years of residency afterwards. And then for some reason I finally picked it up a couple of days ago.

This was a *heck* of a book.

On some level it was as predictable as I expected. It hit all of the Important Internship Tropes, like The Part Where Your Attendings Are Cruel, The Part Where Your Patient Dies Because Of Something You Did, The Part Where You Get Camaraderie With Other Interns, The Part Where You First Realize You Are Actually Slightly Competent At Like One Thing And It Is The Best Feeling In The Universe, The Part Where You Realize How Pointless 99% Of The Medical System Is, The Part Where You Have Sex With Hot Nurses, et cetera.

All I can say is that it was really well done. The whole thing had a touch of magical realism, which turns out to be exactly the right genre for a story about medicine. Real medicine is absolutely magical realist. It's a series of bizarre occurrences just on the edge of plausibility happening to incredibly strange people for life-and-death stakes, day after day after day, all within the context of the weirdest and most byzantine bureaucracy known to humankind.

Just in the past week, for example, I had to deal with an aboulomaniac patient – one with a pathological inability to make up his mind. He came to my clinic for treatment, but as soon as he saw me, he decided he didn't want treatment after all and left. The next day, he was back on my calendar – he'd decided he needed treatment after all – but when his appointment came around, he changed his mind and left again. This happened *five times in five days*. Every day he would phone in asking for an appointment. Every day I would give it to him. Every day he would leave a minute or two before it began. Unsure how to proceed, I sought out my attending. He ignored my questions, pulled me into a side office, took out his cell phone, and started playing me a video. It's a scene from his musical, *The Phantom Of The Psychiatric Unit*, which he's been forcing his interns to rehearse after rounds. I watched, horrified. It was weirdly good.

If I were to write a book about this kind of thing, people would criticize me for being unrealistic. The only way to get away with it is to pass it off as "a touch of magical realism", and this *The House of God* does to excellent effect.

The story revolves around an obvious author-insert character, Roy Basch MD, who starts his internship year at a hospital called the House of God (apparently a fictionalized version of Beth Israel Hospital in Boston). He goes in with

expectations to provide useful medical care to people with serious diseases. Instead, he finds gomers:

"Gomer is an acronym: Get Out of My Emergency Room. It's what you want to say when one's sent in from the nursing home at three A.M."

"I think that's kind of crass," said Potts. "Some of us don't feel that way about old people."

"You think I don't have a grandmother?" asked Fats indignantly. "I do, and she's the cutest dearest, most wonderful old lady. Her matzoh balls float – you have to pin them down to eat them up. Under their force the soup levitates. We eat on ladders, scraping the food off the ceiling. I love..." The Fat Man had to stop, and dabbed the tears from his eyes, and then went on in a soft voice, "I love her very much."

I thought of my grandfather. I loved him too.

"But gomers are not just dear old people," said Fats. "Gomers are human beings who have lost what goes into being human beings. They want to die, and we will not let them. We're cruel to the gomers, by saving them, and they're cruel to us, by fighting tooth and nail against our trying to save them. They hurt us, we hurt them."

This is where the magical realism starts to come in:

Rokitansky was an old bassett. He'd been a college professor and had suffered a severe stroke. He lay on his bed, strapped down, IV's going in, catheter coming out. Motionless, paralyzed, eyes closed, breathing comfortably, perhaps dreaming of a bone, or a boy, or of a boy throwing a bone.

"Mr. Rokitansky, how are you doing?" I asked.

Without opening his eyes, after fifteen seconds, in a husky slurred growl from deep down in his smushed brain he said: PURRTY GUD.

Pleased, I asked, "Mr. Rokitansky, what date is it today?"

PURRTY GUD.

To all my questions, his answer was always the same. I felt sad. A professor, now a vegetable. Again I thought of my grandfather, and got a lump in my throat. Turning to Fats, I said, "This is too sad. He's going to die."

"No, he's not," said Fats. "He wants to, but he won't."

"He can't go on like this."

"Sure he can. Listen, Basch, there are a number of LAWS OF THE HOUSE OF GOD. LAW NUMBER ONE: GOMERS DON'T DIE."

"That's ridiculous. Of course they die."

"I've never seen it, in a whole year here," said Fats.

"They have to."

"They don't. They go on and on. Young people – like you and me – die, but not the gomers. Never seen it. Not once."

"Why not?"

"I don't know. Nobody knows. It's amazing. Maybe they get past it. It's pitiful. The worst."

Potts came in, looking puzzled and concerned. He wanted the Fat Man's help with Ina Goober. They left, and I turned back to Rokitansky. In the dim half-light I thought I saw tears trickling down the old man's cheeks. Shame swept over me. My stomach churned. Had he heard what we'd said?

"Mr. Rokitansky, are you crying?" I asked, and I waited, as the long seconds ticked away, my guilt moaning inside me.

PURRTY GUD.

"But did you hear what we said about gomers?"

PURRTY GUD.

Someone once said that the point of art is to be more real than reality. *The House Of God* is way more real than reality. Reality wishes it could be anywhere close to as real as *The House of God*. This is a world where young people – the kid just out of school, the blushing new mother – die. Even normal old people – your grandmother, your grandpa – can die. But the most decrepit, demented people, the ones for whom every moment of artificially-prolonged life is a gratuitous misery and you pray at every moment that God will just let them find some peace – somehow they never die. They come into the hospital, they go back out to nursing homes, a few weeks later they're back in the hospital, a few weeks later they're back in their nursing homes, but *they never die*. This can't be literally true. But it's the subjective truth of working in a hospital. The Fat Man is right. I've been working in medicine for three years now, and I have seen my share of young people tragically cut off in the prime of life, and yet as far as I can remember I have never seen a gomer die. The magical realism of *House of God* describes the reality of medical professionals infinitely better than the rational world of hospital mortality statistics.

In the world of *The House of God*, the primary form of medical treatment is the TURF – the excuse to get a patient out of your care and on to somebody else's. If the psychiatrist can't stand a certain patient any longer, she finds some trivial abnormality in their bloodwork and TURFs to the medical floor. But she knows that if the medical doctor doesn't want one of *his* patients, then he can interpret a trivial patient comment like "Being sick is so depressing" as suicidal ideation and TURF to psychiatry. At 3 AM on a Friday night, every patient is terrible, the urge to TURF is overwhelming, and a hospital starts to seem like a giant wheel uncoupled from the rest of the world, Psychiatry TURFING to Medicine TURFING to Surgery TURFING to Neurosurgery TURFING to Neurology TURFING back to Psychiatry again. Surely some treatment must get done somewhere? But where? It becomes a legend, The Place Where Treatment Happens, hidden in some far-off hospital wing accessible only to the pure-hearted. This sort of Kafkaesque picture is how medical care feels, and the genius of *The House of God* is that it accentuates the reality just a little bit until its fictional world is almost as magical-realist as the real one.

In the world of *The House of God*, medical intervention can only make patients worse:

Anna O. had started out on Jo's service in perfect electrolyte balance, with each organ system working as

perfectly as an 1878 model could. This, to my mind, included the brain, for wasn't dementia a fail-safe and soothing oblivion of the machine to its own decay?

From being on the verge of a TURF back to the Hebrew House for the Incurables, as Anna knocked around the House of God in the steaming weeks of August, getting a skull film here and an LP there, she got worse, much worse. Given the stress of the dementia work-up, every organ system crumpled: in a domino progression the injection of radioactive dye for her brain scan shut down her kidneys, and the dye study of her kidneys overloaded her heart, and the medication for her heart made her vomit, which altered her electrolyte balance in a life-threatening way, which increased her dementia and shut down her bowel, which made her eligible for the bowel run, the cleanout for which dehydrated her and really shut down her tormented kidneys, which led to infection, the need for dialysis, and big-time complications of these big-time diseases. She and I both became exhausted, and she became very sick. Like the Yellow Man, she went through a phase of convulsing like a hooked tuna, and then went through a phase that was even more awesome, lying in bed deathly still, perhaps dying. I felt sad, for by this time, I liked her. I didn't know what to do. I began to spend a good deal of time sitting with Anna, thinking.

The Fat Man was on call with me every third night as backup resident, and one night, searching for me to go to the ten o'clock meal, he found me with Anna, watching her trying to die.

"What the hell are you doing?" he asked.

I told him.

"Anna was on her way back to the Hebrew House, what happened – wait, don't tell me. Jo decided to go all-out on her dementia, right?"

"Right. She looks like she's going to die."

"The only way she'll die is if you murder her by doing what Jo says."

"Yeah, but how can I do otherwise, with Jo breathing down my neck?"

"Easy. Do nothing with Anna, and hide it from Jo."

"Hide it from Jo?"

"Sure. Continue the work-up in purely imaginary terms, buff the chart with the imaginary results of the imaginary tests, Anna will recover to her demented state, the work-up will show no treatable cause for it, and everybody's happy. Nothing to it."

"I'm not sure it's ethical."

"Is it ethical to murder this sweet gomer with your work-up?"

There was nothing I could say."

After learning these medical secrets, Dr. Basch uses hook and crook to prevent his patients from getting any treatment. They end up healthier than anyone else in the hospital, and Basch becomes a contender for "Most Valuable Intern" – in typical *House of God* style, nobody knows if this award really exists or is just a rumor. His colleagues compete for another award, the "Black Crow", which goes to the intern who gets the most autopsy consents from

grieving families – and which the administration doesn't realize incentivizes doctors to kill their patients. This is so reminiscent of the bizarre incentive systems in real hospitals that it *hurts*.

But as the year goes on, everyone gets more and more frazzled. One intern has a mental breakdown. Another commits suicide by jumping out of a hospital window (this isn't dramatic exaggeration by the way; three junior doctors have committed suicide by jumping out of windows in the past three years in New York City alone). Dr. Basch runs through all sorts of interesting forms of neurosis. Finally, the end of the year approaches, the original crop of interns thinned-out but triumphant – and then they realize they have to do the whole thing again next year as residents, which is maybe a little less grueling but still in the same ballpark.

So they decide, en masse, to go into psychiatry, well-known to be a rare non-terrible residency. The author of *House of God* is a psychiatrist, so I guess this is only a spoiler insofar as you aren't logically omniscient. When the Chief of Medicine learns that every single one of his hospital's interns are going into psychiatry and there aren't going to be any non-psychiatry residents in the whole hospital...

...okay, fine, I won't spoil the ending. But suffice it to say I'm feeling pretty good about my career path right now.

II

House of God does a weird form of figure-ground inversion.

An example of what I mean, taken from politics: some people think of government as another name for the things we do together, like providing food to the hungry, or ensuring that old people have the health care they need. These people know that some politicians are corrupt, and sometimes the money actually goes to whoever's best at demanding pork, and the regulations sometimes favor whichever giant corporation has the best lobbyists. But this is viewed as a weird disease of the body politic, something that can be abstracted away as noise in the system.

And then there are other people who think of government as a giant pork-distribution system, where *obviously* representatives and bureaucrats, incentivized in every way to support the forces that provide them with campaign funding and personal prestige, will take those incentives. *Obviously* they'll use the government to crush their enemies. Sometimes this system also involves the hungry getting food and the elderly getting medical care, as an epiphenomenon of its pork-distribution role, but this isn't particularly important and can be abstracted away as noise.

I think I can go back and forth between these two models when I need to, but it's a weird switch of perspective, where the parts you view as noise in one model resolve into the essence of the other and vice versa.

And *House of God* does this to medicine.

Doctors use certain assumptions, like:

1. The patient wants to get better, but there are scientific limits that usually make this impossible
2. Medical treatment makes people healthier
3. Treatment is determined by medical need and expertise

But in *House of God*, the assumptions get inverted:

1. The patient wants to just die peacefully, but there are bureaucratic limits that usually make this impossible

2. Medical treatment makes people sicker
3. Treatment is determined by what will make doctors look good without having to do much work

Everybody knows that those first three assumptions aren't always true. Yes, sometimes we prolong life in contravention of patients' wishes. Sometimes people mistakenly receive unnecessary treatment that causes complications. And sometimes care suffers because of doctors' scheduling issues. But it's easy to abstract away to an ideal medicine based on benevolence and reason, and then view everything else as rare and unfortunate deviations from the norm. *House of God* goes the whole way and does a full figure-ground inversion. The outliers become the norm; good care becomes the rare deviation. What's horrifying is how convincing it is. Real medicine looks *at least* as much like the bizarro-world of *House of God* as it does the world of the popular imagination where doctors are always wise, diagnoses always correct, and patients always grateful.

There have been a couple of studies finding that giving people health insurance doesn't make them any healthier – see for example the [RAND Health Insurance Experiment](#) and the [Oregon Medicaid Experiment](#). I've always been skeptical of these studies, because it seems logical that people who can afford health care will get more of it, and there are ten zillion studies showing various forms of health care to help. Insulin helps diabetes. Antibiotics help sepsis. Surgery helps appendicitis. To deny claims like these would be madness, yet the studies don't lie. What is going on?

And the answer has to be somewhere in the bizarro-world of *House of God*. Real medical treatment looks precious little like the House MD model of rare serious disease → diagnosis → cure. At least as often, it's like the *House of God* model where someone becomes inconvenient → send to hospital → one million unnecessary tests. Everyone agrees this is part of the story. *House of God* is a brilliant book in that it refactors perception to place it in the foreground.

But it's brilliant because in the end it's *not* just a romp through hilarious bureaucratic mishaps. There is as much genuine human goodness and compassion in this book as there is in any rousing speech by a medical school dean. The goodness is often mixed with horror – the doctor who has to fight off hordes of autopsy-consent-form-seekers to let a dying patient spend his last few seconds in peace, or the one who secretly slips euthanasia to a terminal patient begging for an end to the pain because he knows it's the right thing to do.

The question posed here is “what do you do in a crazy cannibalistic system where it's impossible to do good work and everyone is dying all around you?”, and the answer is “try as hard as you can to preserve whatever virtue you can, and to remain compassionate and human”. The protagonist swings wildly between “this is all bullshit and I'll just make fun of these disgusting old people and call it a day” and “I need to save everybody and if I don't I should hate myself forever”, and eventually like everybody, comes to some kind of synthesis where he recognizes he's human, recognizes that his patients are human, and tries to deal with it with whatever humor and grace he can manage.

It's hard enough for a book to be funny, and it's hard enough for one to be deep, but a book like *House of God* that can be both at once within the space of a few sentences is an absolute treasure.

III

I talked to my father about *House of God*, and I told him a few parts that seemed unrealistic. He told me that those parts were 100% true in 1978 when the book was written. I looked into it more, and ended up appreciating the work on a whole new level. *Uncle Tom's Cabin* is credited with kickstarting the emancipationist movement and maybe even [causing the Civil War](#). *The Jungle* is famous for launching a whole new era of safety regulations. *House of God* has a place beside them in the pantheon of books that have changed the world.

The book's “Second Law” is “GOMER GOES TO GROUND”: demented old people will inevitably fall out of their hospital

bed and injure themselves. The book has a whole funny/horrifying scene where the senior resident explains his strategy for this eventuality: He leaves their beds low enough that patients won't kill themselves when they fall, but high enough that they'll probably break a bone or two and have to go to orthopaedic surgery – which takes them off his hands. Later, a medical student apes this procedure, a patient falls and breaks a bone or two, and everyone freaks out and tells him that it was a joke, that *of course* you don't really arrange skeletal fractures for old people just to save yourself time, what kind of heartless moron could *think* such a thing? This is some nth-level meta-humor: the reader probably mistook it for real advice because it meshes so seamlessly with all of the other madness and horror, yet most of the other madness and horror in the book is easily recognizable by practicing doctors as a real part of the medical system. Actually, on the n+1st meta-level, I'm not at all sure that the resident wasn't meant to be completely serious and then backtracked and called it a joke when it went wrong. For that matter, I'm far from sure this wasn't a real medical practice in the 1970s.

I see enough falls that I wasn't surprised to see them as a theme, but I thought the book exaggerated their omnipresence. My father said it didn't – there were just far more falls back in the Old Days. Now hospitals are safer and falls are comparatively rare. Why? Because the government passed a law saying that insurance wouldn't pay hospitals extra money for the extra days patients have to stay due to fall-related injuries. I am so serious about this. This, I think, is the n+2nd meta-level; amidst all its jokes-played-straight the book treats encouraging falls as an actual in-universe joke, and yet in the real world once hospitals were no longer incentivized to let patients fall the falls stopped.

How did people become aware of this kind of thing? How did the movement against it start? A lot of it seems to be because of *House of God*. Everyone in medicine knew about this sort of thing. But *House of God* made it common knowledge.

People were scared to speak up. Everyone thought that maybe they were just a uniquely bad person, or their hospital a uniquely bad institution. Anyone who raised some of these points was met with scorn by prestigious doctors who said that maybe they just weren't cut out for medicine. *House of God* shaped medicine because it was the first thing to say what everybody was experiencing. Its terms like "gomer" and "turf" made it into the medical lexicon because they pointed to obvious features of reality nobody had the guts to talk about before.

Shem writes an afterword where he talks about the reaction to the book. Junior doctors and the public loved it. Senior doctors hated it. He tells the story of going to a medical conference. Someone asked who he was, and he said jokingly "I'm the most hated doctor here". His interlocutor answered "Oh, don't worry, I'm sure you're not as bad as the guy who wrote that *House of God* book."

But *House of God* gets credit for helping start movements to cut intern work hours, protect doctors from sleep deprivation, reduce patient falls, and teach empathy and communication skills. The moral of the story is: the courage to tell the truth is rare and powerful. More specifically: the courage to tell the truth is rare and powerful not just in Stalinist dictatorships and violent cults, but in apparently normal parts of everyday First World life. All of these differently loaded terms like "culture of silence" and "political correctness" point at a fear of rocking various boats with nothing but your imperfect first-person knowledge to go on. But a tiny crack in the wall can make a big difference.

IV

In a closing scene, Dr. Basch and all of his fellow interns – interns who had broken into tears weekly, gotten burnt out, starting seeing psychiatrists, considered suicide, all this stuff, these interns who had smashed up against the unendurable horrors of medicine and held themselves together only by the promise that it would soon be over – the minute they graduate internship they change their tune:

It looked like all but two or three [interns] would stay. The Runt and I were definitely leaving; Chuck hadn't yet said. The others were staying. In years to come they would spread out across America into academic centers and Fellowships, real red-hots in internal medicine, for they had been trained at the Best Medical School's best House, the House of God. Although a few might kill themselves or get addicted or go crazy, by and large they'd repress and conform and perpetuate the Leggo [the Chief of Medicine] and the House and all the best medical stuff. [Eddie] had been praised by the Leggo that he could start off the second year as ward resident, with "a free rein" on his interns. And so, saying already that the internship been "not so bad," he was preparing to indoctrinate his new charges: "I want them on their knees from day one."

Shem's author mouthpiece character Berry says:

It's been inhuman. No wonder doctors are so distant in the face of the most poignant human dramas. The tragedy isn't the crassness, but the lack of depth. Most people have some human reaction to their daily work, but doctors don't. It's an incredible paradox that being a doctor is so degrading and yet is so valued by society. In any community, the most respected group are doctors. [It's] a terrific repression that makes doctors really believe that they are omnipotent healers. If you hear yourselves saying, 'Well, this year wasn't really that bad,' you're repressing, to put the next group through it. [But] it's hard to say no. If you're programmed from age six to be a doctor, invest years in it, develop your repressive skills so that you can't even recall how miserable you were during internship, you can't stop.

Shem's thesis is that it isn't just about not wanting to make waves or offend the Chief of Medicine. It's about denying your own pain by identifying with the system.

This puts me in a weird spot. My internship (I find myself saying) wasn't so bad. I can give you some arguments why this might be true – things have gotten a lot better since *The House of God* was published (with no small credit to Shem himself), a small community hospital in Michigan is less intense than Harvard Medical School's training hospital, psychiatry interns sometimes have it easier than internal medicine interns since everyone knows this isn't a permanent deal for them.

And yet I distinctly remember one night a long time ago, coming home from high school. I had noticed that all of the adults around me said high school was some of the best years of their lives and I would miss it when I was gone, and yet high school seemed objectively terrible. I wondered if there might be some bias or bizarre shift in memory that happened sometime in people's twenties and gave them a localized amnesia or insanity. So I very distinctly recall telling myself "My current assessment is that high school is terrible, and if you ever find yourself remembering that high school was lovely, please be aware that your memories have been hijacked by some malevolent force."

And God help me, but *every single part of my brain is telling me that high school was lovely*. I fondly remember all the friends I made, the crazy teachers I had to put up with, the science competitions I won, the lunches spent in the library reading whatever random stuff I could get my hands on. It seems like it was a blast. It's hard for me to even trust that one memory as anything more than imagination or the product of a single bad day. But although high-school-me had a lot of issues, he generally had a decent head on his shoulders, and if he says my memories have been hijacked, then I grudgingly believe him.

So was my intern year a good learning experience? *I have no idea and I'm not sure anyone else does either*. It's another type of figure-ground inversion: parade of [horrors](#) broken only by the occasional triumph, or clear sailing with a few bad moments?

On my last day of internship, one of my colleagues who was moving on said "I'm going to miss hating this place". I've always remembered that phrase. Now I wonder if it's some kind of weird snapshot of the exact moment of transition,

the instant when “nightmarish ordeal” morphs into “halcyon days of youth”. This is why medicine has to be written as magical realism. How else to capture a world where people reliably go from agony to Stockholm Syndrome in the space of a day, and where the transition is so intermixed with the general weirdness that it doesn’t even merit special remark?

I found myself having more emotions reading *House of God* than I’ve had about anything in a long time. I don’t really know why. But I think it has something to do with this resignation to the general incommunicable weirdness all around anyone who works in medicine. Somehow Shem manages to avoid the normalization of insanity that happens to every young doctor, capture the exact subjective experience and write it down in a way that makes sense. And then, having put his finger right on the unbearable thing, he makes it funny and beautiful and poignant.

I tell her. Again I tell her about Dr. Sanders bleeding out in my lap, about the look in Potts’s eyes that night before he jumped, about my pushing the KCl into poor Saul. I tell her how ashamed I am for turning into a sarcastic bastard who calls the old ones gomers, how, during the ternship, I’d ridiculed them for their weaknesses, for throwing up their suffering in my face, for scaring me, for forcing me to do disgusting things to take care of them. I tell her how I want to live, compassionately, with the idea of death clearly in sight, and how I doubt I can do that, ever again. As I think back to what I’d gone through and what I’d become, sadness wells up and mixes with contempt. I put my head into Berry’s folds and weep, and curse, and shout, and weep.

“... and in your own way, you did. Someone had to care for the gomers; and this year, in your own way, you did.”

“The worst thing is this bitterness. I used to be different, gentle, even generous, didn’t I? I wasn’t always like this, was I?”

“I love who you are. To me, underneath it all, you’re still there.” She paused, and then, eyes sparkling, said, “And you might even be better.”

“What? What do you mean?”

“This might have been the only thing that could have awakened you. Your whole life has been a growing from the outside, mastering the challenges that others have set for you. Now, finally, you might just be growing from inside yourself.

He also frames all of it in the language of psychoanalysis, which is jarring and sounds preachy. I’ve ordered the sequel, *Mount Misery*, about his training as a psychoanalyst. Expect a review of that soon.

Side Effects May Include Anything

Posted on January 20, 2016 by Scott Alexander



A couple of days ago a patient said he'd become depressed after starting Xolair, a new asthma drug I know nothing about.

On the one hand, lots of things that mess with the immune system can cause depression. On the other, patients are notorious for blaming drugs for any random thing that happens around the same time they started taking them. So I did what any highly-trained competent medical professional would: I typed "does xolair cause depression?" into Google.

The results seemed promising. The first site was called "Can Xolair cause depression?". The second was "Is depression a side effect of Xolair?". Also on the front page were "Could Xolair cause major depression?" and "Xolair depression side effects". Clearly this is a well-researched topic that lots of people cared about, right?

Let's look closer at one of those sites, EHealthMe.com. It [says](#): "Major depression is found among people who take Xolair, especially for people who are female, 40-49 old, also take medication Singulair, and have Asthma. We study 11,502 people who have side effects while taking Xolair from FDA and social media. Among them, 14 have Major depression. Find out below who they are, when they have Major depression and more." Then it offers a link: "Join a support group for people who take Xolair and have Major depression".

First things first: if there were actually 11502 people taking Xolair, and only 14 of them had major depression, that would be a rate of 0.1%, compared to 6.9% in the general population. In other words, Xolair would be the most effective antidepressant on Earth. But of course nobody has ever done an n=11502 study on whether a random asthma medication causes depression, and EHealthMe is just scraping the FDA databases to see how many people reported depression as a side effect to the FDA. But only a tiny percent of people who get depression report it, and depression sometimes strikes at random times whether you're taking Xolair or not. So this tells us nothing.

And yet a patient who worries that Xolair might be causing their depression will Google "can xolair cause depression?", and she will end up on this site that says "major depression is found among people who take Xolair", which is one of the worst examples of weasel words I've ever heard. Then she will read that there are *entire support groups* for depressed Xolair sufferers. She will find all sorts of scary-looking information like that Xolair-related depression has been increasing since 2008. And this is above and beyond just the implications of somebody bothering to write an entire report about the Xolair-depression connection!

In case you haven't guessed the twist – no one's ever investigated whether Xolair causes depression. EHealthMe's business model is to make an automated program that runs through every single drug and every possible side effect, scrapes the FDA database for examples, then autopublishes an ad-filled web page titled "COULD \$DRUG CAUSE \$SIDE_EFFECT?". It populates the page by spewing random FDA data all over it, concludes "\$SIDE_EFFECT is found among people who take \$DRUG", and offers a link to a support group for \$DRUG patients suffering from \$SIDE_EFFECT. Needless to say, the support group is an automatically-generated forum with no posts in it.

And it's not just EHealthMe. This is a whole market, with competitors elbowing their way past one another to the top of the Google search results. Somebody who doubts EHealthMe and seeks an online second opinion will probably just end up at PatientsVille, [whose page](#) is called "Xolair Depression Side Effects", which contains the same FDA data, and

which gets the Google description text “This opens a possibility that Xolair could cause Depression”. Or Treato, [whose page](#) claims to contain 56 reader comments on Xolair and depression, but which has actually just searched the Web for every single paragraph that contains “Xolair” and “depression” together and then posted garbled excerpts in its comment section. For example, one of their comments – and this is not at all clear from Treato’s garbled excerpt – is from [a tennis forum](#), where a user with the handle Xolair talks about how his tennis serve is getting worse with age; another user replies “Xolair, I read this and get depressed, I just turned 49.” But if you don’t check whether it came from a tennis forum or not, 56 reports of a connection between a drug and a side effect sounds convincing!

This is *really scummy*. Maybe it’s not the most devious of traps for you or me, but what about for your grandmother? What about for those people who send money to Nigerian princes? The law is usually pretty strict about who can and can’t provide medical information – so much so that it [cracks down on 23andMe](#) just for reading off the genome in a way that uneducated people might misinterpret. Yet somehow sites like EHealthMe are allowed to continue, because they just *very strongly imply* fake medical information instead of saying it outright.

Remember, [only about 50% of people](#) who are prescribed medication take it. Sometimes it’s personal choice or simple forgetfulness. But a lot of the time they stop because of side effects. I had a patient a few months ago who was really depressed. I started her on an antidepressant and she got much better. Then she stopped the medication cold turkey and got a lot worse again. I asked her why she’d stopped. She said her shoulder started hurting, she’d Googled whether antidepressants could cause shoulder pain, and read that they could. She couldn’t remember what site she was reading, but I bet it was EHealthMe or Treato or some of the others just like them.

One day, somebody’s going to Google “can penicillin cause cancer?”, read a report with a link to a support group for penicillin-induced-cancer survivors, stop taking antibiotics, and die. And when that happens, I hope it’s in America, so I can be sure their family will sue the company involved for more money than exists in the entire world.

Cost Disease In Medicine: The Practical Perspective

Posted on June 20, 2018 by Scott Alexander



Sometimes I imagine quitting my job and declaring war on cost disease in medicine.

I would set up a practice with a name like Cheap-O Psychiatry. The corny name would be important. It would be a statement of values. It would weed out the people who would say things like "How dare you try to put a dollar value on the health of a human being!" Those people are how we got into this mess, and they would be welcome to keep dealing with the unaffordable health system they helped create. Cheap-O Psychiatry would be for everyone else.

Cheap-O Psychiatry wouldn't have an office, because offices cost money. You would Skype, from your house to mine. It wouldn't have a receptionist, because receptionists cost money. You would book a slot in my Google Calendar. It wouldn't have a billing department, because billing departments cost money. You would PayPal me the cost of the appointment afterwards – or, to be really #aesthetic, use cryptocurrency.

The Cheap-O website would include a library of great resources on every subject. How To Eat Right. How To Get Good Sleep. How To Find A Good Therapist. The Cognitive Behavioral Therapy Workbook. The Meditation Relaxation Tape. But the flip side would be that Cheap-O appointments would be brutally efficient. If you had problems with sleep, I would evaluate you for any relevant diseases, give you any medications that might be indicated, then tell you to read the How To Get Good Sleep guide on the website. Boom, done. Small talk would be absolutely banned.

How little could Cheap-O charge? Suppose I wanted to earn an average psychiatrist salary of about \$200K – the whole point of cost disease is that we should be able to lower prices without anyone having to take a pay cut. And suppose I work a 40 hour week, 50 weeks a year, each appointment takes 15 minutes, and 75% of my workday is patient appointments. That's 6000 appointments per year. So to make my \$200K I would need to charge about \$35 per appointment. There would be a few added costs – malpractice insurance would probably run about \$10K per year – but this is the best-case scenario.

\$35 per appointment isn't bad. Most existing cash-only psychiatry practices charge at least \$150 per (thirty minute) appointment, so we would be less than a quarter of the going rate. I think a lot of insurances charge a \$40 copay per psychiatrist visit, so even *uninsured* Cheap-O patients would be paying less cash than *insured* patients anywhere else. Create Cheap-O style psychiatry offices, primary care offices, etc, all around the country, and maybe (aside from catastrophe insurance, which should be cheap) having health insurance would no longer be such a big deal.

My job is great and I love it, so I'm only slightly tempted to do this myself. The reason I bring it up is: why doesn't anyone else do it? And if it's possible to provide cheap health care like this, then how does health care still cost so much? What am I missing?

I don't know. I've never run a business and it's probably much harder and more expensive than I think. One of the reasons for my Cheap-O fantasy is so that I could find out. But here are some speculations.

Part of the reason might be because there's a shortage of doctor-entrepreneurs, and the few existing doctor-entrepreneurs are busy finding new ways to make ultra-boutique-super-premium clinics that they charge rich people \$500/hour for the privilege of entering. And they're doing that because it pays way more than \$200K/year. \$200K/year is the standard salary for an average psychiatrist who wants a zero-risk job in the current system, and the

privilege of never having to worry about the business side of things.

Another part might be that insurance is squatting in the mid-range market. Even in America, most people are insured. So unless your cost can beat the insurance co-pay – which even Cheap-O barely does under ideal conditions – most people will go to standard insurance-accepting practices unless you give them a good reason not to. And the best reason not to will be that you’re claiming to be better than insurance-accepting clinics – which means you’re aiming at the high-end market. And insurance-accepting practices can’t lower prices because insurances make you follow lots of rules before they’ll work with you, plus you need a small city worth of administrators to deal with the insurance companies.

(also, seeing a patient every fifteen minutes is exhausting; one of the advantages of hour-long appointments is that most people don’t need an hour and so you can take the last twenty minutes to write notes or answer messages or work on blog posts)

Right now the only way I can imagine the niche getting filled is somebody doing it ~~for the lulz~~ as an act of political protest. Imagine if someone started Cheap-O Psychiatry and it *worked*. All of this stuff I’m saying about how socialized health care might be better than our current system but isn’t the real answer, how it’s just locking in entrenched cost disease and a truly free market could find better alternatives – instead of vaguely gesturing at it, there would finally be some evidence.

(Well, there’s already the Surgery Center of Oklahoma, which [does exactly this](#) and costs about a fifth as much as surgery anywhere else. But maybe if there’s *more* evidence, people will stop ignoring it.)

The Amish Health Care System

Posted on April 20, 2020 by Scott Alexander



I

Amish people spend only a fifth as much as you do on health care, and their health is fine. What can we learn from them?

A reminder: the Amish are a German religious sect who immigrated to colonial America. Most of them live apart from ordinary Americans (who they call “the English”) in rural communities in Pennsylvania and Ohio. They’re famous for their low-tech way of life, generally avoiding anything invented after the 1700s. But this isn’t absolute; they are willing to accept technology they see as a net positive. Modern medicine is in this category. When the Amish get seriously ill, they will go to modern doctors and accept modern treatments.

The Muslims claim Mohammed was the last of the prophets, and that after his death God stopped advising earthly religions. But sometimes modern faiths will make a decision so inspired that it could only have come from divine revelation. This is how I feel about the Amish belief that health insurance companies are evil, and that good Christians must have no traffic with them.

And Deists believe that God is like a watchmaker, an artisan who built the world but does not act upon it. But by some miracle, the US government played along and granted the Amish exemptions from all the usual health care laws. They don’t have to pay Medicare taxes or social security. They aren’t included in the Obamacare mandate. They can share health care costs the way they want, ignoring any regulations to the contrary. They are genuinely on their own.

They’ve ended up with a simple system based on church aid. Everyone pays tithes to their congregation (though they don’t call it that). The churches meet in houses and have volunteer leaders, so expenses minimal. Most of the money goes to “alms” which the bishop distributes to members in need. This replaces the social safety net, including health insurance. Most Amish go their entire life without needing anything else.

About a third of Amish are part of a more formal insurance-like institution called Amish Hospital Aid. Individuals and families pay a fixed fee to the organization, which is not-for-profit and run by an unpaid board of all-male elders. If they need hospital care, AHA will pay for it. How does this interact with the church-based system? [Rohrer and Dundes](#), my source for most of this post, say that it’s mostly better-off Amish who use AHA. Their wealth is tied up in their farmland, so it’s not like they can use it to pay hospital bills. But they would feel guilty asking their church to give them alms meant for the poor. AHA helps protect their dignity and keep church funds for those who need them most.

How well does this system work?

The Amish outperform the English [on every measured health outcome](#). 65% of Amish rate their health as excellent or very good, compared to 58% of English. Diabetes rates are 2% vs. 8%, heart attack rates are 1% vs. 6%, high blood pressure is 11% vs. 31%. Amish people go to the hospital about a quarter as often as English people, and this difference is consistent across [various categories of illness](#) (the big exception is pregnancy-related issues – most Amish women have five to ten children). This is noticeable enough that lots of health magazines have articles on [The Health Secrets of the Amish](#) and [Amish Secrets That Will Add Years To Your Life](#). As far as I can tell, most of the secret

is spending your whole life outside doing strenuous agricultural labor, plus being at a tech level two centuries too early for fast food.

But Amish people also die earlier. Lots of old studies say the opposite – for example, [this one](#) finds Amish people live longer than matched Framingham Heart Study participants. But things have changed since Framingham. The Amish have had a life expectancy in the low 70s since colonial times, when the rest of us were dying at 40 or 50. Since then, Amish life expectancy has stayed the same, and English life expectancy has improved to the high 70s. The most recent Amish estimates I have still say [low seventies](#), so I think we are beating them now.

If they're healthier, why is their life expectancy lower? Possibly they are less interested in [prolonging life](#) than we are. R&D write:

Amish people are more willing to stop interventions earlier and resist invasive therapies than the general population because, while they long for healing, they also have a profound respect for God's will. This means taking modest steps toward healing sick bodies, giving preference to natural remedies, setting common-sense limits, and believing that in the end their bodies are in God's hands.

The Amish health care system has an easier job than ours does. It has to take care of people who are generally healthy and less interested in extreme end-of-life care. It also supports a younger population – because Amish families have five to ten children, the demographics are weighted to younger people. All of these make its job a little bit simpler, and we should keep that in mind for the following sections.

How much do the Amish pay for health care? This is easy to answer for Amish Hospital Aid, much harder for the church system.

Amish Hospital Aid charges \$125 monthly per individual or \$250 monthly per family (remember, Amish families can easily be ten people). Average US health insurance costs \$411 monthly per individual (Obamacare policies) or \$558 monthly per individual (employer sponsored plan; employers pay most of this). I'm not going to bother comparing family plans because the definition of "family" matters a lot here. On the surface, it looks like the English spend about 4x as much as the Amish do.

But US plans include many more services than AHA, which covers catastrophic hospital admissions only. The government bans [most Americans](#) from buying plans like this; they believe it's not enough to count as real coverage. The cheapest legal US health plan varies by age and location, but when I take my real age and pretend that I live near Amish country, the government offers me a \$219/month policy on Obamacare. This is only a little higher than what the Amish get, and probably includes more services. So here it seems like the Amish don't have much of an efficiency advantage. They just make a different tradeoff. It's probably the right tradeoff for them, given their healthier lifestyle.

But remember, only a third of Amish use AHA. The rest use a church-based system? How does that come out?

It's hard to tell. Nobody agrees on how much Amish tithe their churches, maybe because different Amish churches have different practices. R&D suggest families tithe 10% of income, [this article](#) on church-based insurances says a flat \$100/month fee, and this "[Ask The Amish](#)" column says that churches have twice-yearly occasions where they ask for donations in secret and nobody is obligated to give any particular amount ("often husbands and wives won't even know how much the other is giving.") So it's a mess, and even knowing the exact per-Amish donation wouldn't help, because church alms cover not just health insurance but the entire social safety net; the amount that goes to health care probably varies by congregation and circumstance.

A few people try to estimate Amish health spending directly. [This ABC story](#) says \$5 million total for all 30,000 Amish in Lancaster County, but they give no source, and it's absurdly low. [This QZ story](#) quotes Amish health elder Marvin

Wengerd as saying \$20 – \$30 million total for Lancaster County, which would suggest health spending of between \$600-\$1000 per person. This sounds potentially in keeping with some of the other estimates. A \$100 per month tithe would be \$1200 per year – if half of that goes to non-health social services, that implies \$600 for health. The average Amish family earns about \$50K (the same as the average English family, somehow!) so a 10% tithe would be \$5000 per year, but since the average Amish family size is seven children, that comes out to about \$600 per person again. So several estimates seem to agree on between \$600 and \$1000 per person.

One possible issue with this number: does Wengerd know how much Amish spend out of pocket? Or does his number just represent the amount that the official communal Amish health system spends? I'm not sure, but taking his words literally it's total Amish spending, so I am going to assume it's the intended meaning. And since the Amish rarely see doctors for minor things, probably their communal spending is a big chunk of their total.

Update: an SSC reader is able to contact his brother, a Mennonite deacon, for better numbers. [He says that](#) their church spends an average of \$2000 per person (including out of pocket).]

How does this compare to the US as a whole? The [National Center For Health Statistics](#) says that the average American spends \$11,000 on health care. This suggests that the average American spends between five and ten times more on health care than the average Amish person.

How do the Amish keep costs so low? R&D (plus a few other sources) identify some key strategies.

First, the Amish community bargains collectively with providers to keep prices low. This isn't unusual – your insurance company does the same – but it nets them better prices than you would get if you tried to pay out of pocket at your local hospital. [This article](#) gives some examples of Amish getting sticker prices discounted from between 50% to 66% with this tactic alone; Medicare gets about the same.

Second, the Amish are honorable customers. This separates them from insurance companies, who are constantly trying to scam providers however they can. Much of the increase in health care costs is "administrative expenses", and much of these administrative expenses is hiring an army of lawyers, clerks, and billing professionals to thwart insurance companies' attempts to cheat their way out of paying. If you are an honorable Amish person and the hospital knows you will pay your bill on time with zero fuss, they can waive all this.

But can this really be the reason Amish healthcare is cheaper? When insurance companies negotiate with providers, patients are on the side of the insurances; when insurance companies get good deals (eg a deal of zero dollars because the insurance has scammed the hospital), the patient's care is cheaper, and the insurance company can pass some of those savings down as lower prices. If occasionally scamming providers meant insurance companies had to pay more money total, then they would stop doing it. My impression is that the real losers here are uninsured patients; absent any pressure to do otherwise, hospitals will charge them the sticker price, which includes the dealing-with-insurance-scams fee. The Amish successfully pressure them to waive that fee, which gets them better prices than the average uninsured patient, but still doesn't land them ahead of insured people.

Third, Amish don't go to the doctor for little things. They either use [folk medicine](#) or [chiropractors](#). Some of the folk medicine [probably works](#). The chiropractors probably don't, but they play a helpful role reassuring people and giving them the appropriate obvious advice while telling the really serious cases to seek outside care. With this help, Amish people mostly avoid primary care doctors. [Holmes County health statistics](#) find that only 16% of Amish have seen a doctor in the past year, compared to 54% of English.

Fourth, the Amish [never sue doctors](#). Doctors around Amish country know this, and give them the medically indicated level of care instead of practicing "defensive medicine". If Amish people ask their doctors to be financially considerate – for example, let them leave the hospital a little early – their doctors will usually say yes, whereas *your doctor would*

say no because you could sue them if anything went wrong. In some cases, Amish elders formally promise that no member of their congregation will ever launch a malpractice lawsuit.

Fifth, the Amish don't make a profit. Church aid is dispensed by ministers and bishops. Even Amish Hospital Aid is run by a volunteer board. None of these people draw a salary or take a cut. I don't want to overemphasize this one – people constantly obsess over insurance company profits and attribute all health care pathologies to them, whereas in fact they're a low single-digit percent of costs (did you know Kaiser Permanente is a nonprofit? Hard to tell, isn't it?) But every little bit adds up, and this is one bit.

Sixth, the Amish don't have administrative expenses. Since the minister knows and trusts everyone in his congregation, the "approval process" is just telling your minister what the problem is, and the minister agreeing that's a problem and giving you money to solve it. This sidesteps a lot of horrible algorithms and review boards and appeal boards and lawyers. I don't want to overemphasize this one either – insurance companies are legally required to keep administrative expenses low, and most of them succeed. But again, it all adds up.

Seventh, the Amish feel pressure to avoid taking risks with their health. If you live in a tiny community with the people who are your health insurance support system, you're going to feel awkward smoking or drinking too much. Realistically this probably blends into a general insistence on godly living, but the health insurance aspect doesn't hurt. And I'm talking like this is just informal pressure, but occasionally it can get very real. R&D discuss the case of some Amish teens who get injured riding a snowmobile – forbidden technology. Their church decided this was not the sort of problem that godly people would have gotten themselves into, and refused to help – their families were on the hook for the whole bill.

Eighth, for the same reason, Amish try not to overspend on health care. I realize this sounds insulting – other Americans aren't trying? I think this is harsh but true. Lots of Americans get an insurance plan from their employer, and then consume health services in a price-insensitive way, knowing very well that their insurance will pay for it. Sometimes they will briefly be limited by deductibles or out-of-pocket charges, but after these are used up, they'll go crazy. You wouldn't believe how many patients I see who say things like "I've covered my deductible for the year, so you might as well give me the most expensive thing you've got", or "I'm actually feeling fine, but let's have another appointment next week because I like talking to you and my out-of-pocket charges are low."

But it's not just avoiding the obvious failure modes. Careful price-shopping can look very different from regular medical consumption. Several of the articles I read talked about Amish families traveling from Pennsylvania to Tijuana for medical treatment. One writer [describes](#) Tijuana clinics sending salespeople up to Amish Country to advertise their latest prices and services. For people who rarely leave their hometown and avoid modern technology, a train trip to Mexico must be a scary experience. But prices in Mexico are cheap enough to make it worthwhile.

Meanwhile, back in the modern world, I've [written before](#) about how a pharma company took clonidine, a workhorse older drug that costs \$4.84 a month, transformed it into Lucemyra, a basically identical drug that costs \$1,974.78 a month, then created a rebate plan so that patients wouldn't have to pay any extra out-of-pocket. Then they told patients to ask their doctors for Lucemyra because it was newer and cooler. Patients sometimes went along with this, being indifferent between spending \$4 of someone else's money or \$2000 of someone else's money. Everything in the US health system is like this, and the Amish avoid all of it. They have a normal free market in medical care where people pay for a product with their own money (or their community's money) and have incentives to check how much it costs before they buy it. I *do* want to over-emphasize this one, and honestly I am surprised Amish health care costs are *only* ten times cheaper than ours are.

I don't know how important each of these factors is, or how they compare to more structural factors like younger populations, healthier lifestyles, and less end-of-life care. But taken together, they make it possible for the Amish to get health care without undue financial burden or government support.

II

Why look into the Amish health system?

I'm fascinated by how many of today's biggest economic problems just mysteriously failed to exist in the past. Our grandparents easily paid for college with summer jobs, raised three or four kids on a single income, and bought houses in their 20s or 30s and never worried about rent or eviction again. And yes, they got medical care without health insurance, and avoided the kind of medical bankruptcies we see too frequently today. How did this work so well? Are there ways to make it work today? The Amish are an extreme example of people who try to make traditional systems work in the modern world, which makes them a natural laboratory for this kind of question.

The Amish system seems to work well for the Amish. It's hard to say this with confidence because of all the uncertainties. The Amish skew much younger than the "English", and live much healthier lifestyles. Although a few vague estimates suggest health care spending far below the English average, they could be missing lots of under-the-table transactions. And again, I don't want to ignore the fact that the Amish do live a little bit shorter lives. You could tell a story where all of these add up to explain 100% of the difference, and the Amish aren't any more efficient in their spending at all. I don't think this is right. I think the apparent 5x advantage, or something like it, is real. But right now this is just a guess, not a hard number.

What if it is? It's hard to figure out exactly what it would take to apply the same principles to English society. Only about a quarter of Americans attend church regularly, so church-based aid is out. In theory, health insurance companies ought to fill the same niche, with maybe a 10% cost increase for profits and overhead. Instead we have a 1000% cost increase. Why?

Above, I said that the most important factor is that the Amish comparison shop. Everyone needs to use other people's money to afford expensive procedures. But for the Amish, those other people are their fellow church members and they feel an obligation to spend it wisely. For the English, the "other people" are faceless insurance companies, and we treat people who don't extract as much money as possible from them as insufficiently savvy. But there's no easy way to solve this in an atomized system. If you don't have a set of thirty close friends you can turn to for financial help, then the only institutions with enough coordination power to make risk pooling work are companies and the government. And they have no way of keeping you honest except the with byzantine rules about "prior authorizations" and "preferred alternatives" we've become all too familiar with.

(and as bad as these are, there's something to be said for a faceless but impartial bureaucracy, compared to having all your neighbors judging your lifestyle all the time.)

This is a neat story, but I have two concerns about it.

First, when I think in terms of individual people I know who have had trouble paying for health care, it's hard for me to imagine the Amish system working very well for them. Many have chronic diseases. Some have mysterious pain that they couldn't identify for years before finally getting diagnosed with something obscure. Amish Hospital Aid's catastrophic policy would be useless for this, and I feel like your fellow church members would get tired of you pretty quickly. I'm not sure how the Amish cope with this kind of thing, and maybe their system relies on a very low rate of mental illness and chronic disease. A lot of the original "hygiene hypothesis" work was done on the Amish, their autoimmune disease rates are amazing, and when you take out the stresses of modern life maybe a lot of the ailments the American system was set up to deal with just stop being problems. I guess my point is that the numbers seem to work out, and the Amish apparently remain alive, but when I imagine trying to apply the Amish system to real people, even assuming those real people have cooperative churches and all the other elements I've talked about, I can't imagine it doing anything other than crashing and burning.

Second, I don't think this is *actually* how our grandparents did things. I asked my literal grandmother, a 95 year old former nurse, how health care worked in her day. She said it just wasn't a problem. Hospitals were supported by wealthy philanthropists and religious organizations. Poor people got treated for free. Middle class people paid as much as they could afford, which was often the whole bill, because bills were cheap. Rich people paid extra for fancy hospital suites and helped subsidize everyone else. Although most people went to church or synagogue, there wasn't the same kind of Amish-style risk pooling.

This makes me think that the Amish method, even though it works, isn't the method that worked for past generations. It's an innovation intended to cover for health care prices being higher than anything that traditional societies had to deal with.

Why did health care prices start rising? I've wondered about this a lot before – see [here](#), [here](#), and [here](#). Looking into this issue, I noticed glimpses of a different possibility. The increase started around the same time that health insurance began to spread. In one sense, this is unsurprising – of course health insurance would become a thing around the time care became unaffordable. But I've never seen someone really try to tease out causality here. Might the two trends have been mutually self-reinforcing? The price of care rises due to some original shock. Someone invents health insurance, which seems like a good idea. But this creates a series of perverse incentives, which other actors figure out how to exploit (eg the Lucemyra example above). Insurance-based-health-care becomes less efficient, but hospitals can't or don't internalize this to the insured patients – they just raise the price for everyone, insurance or no. That makes even more people need health insurance, and the cycle repeats as prices grow higher and higher and insurance becomes more and more necessary. This syncs well with some explanations I've heard of rising college prices, where once the government made easy loans and subsidies available to everyone, prices rose until they consumed all the resources available.

I have no idea if this is true or not. If it is, the Amish succeed partly by successfully forcing providers to internalize the costs of insurance to insurance patients. Sometimes they do this by literally asking hospitals for better prices because they are not insured (eg the "honest customer" example above). Other times they flee the country entirely to reach a medical system that doesn't deal with insured patients (eg Tijuana). This seems to work well for them. But their reliance on church alms and Amish Hospital Aid suggests that their care is still more expensive and burdensome for them than past generations' care was for them. They've just learned ways to manage the expense successfully.

Sociology and Economics



“Well, in our country,” said Alice, still panting a little, “you’d generally get to somewhere else—if you run very fast for a long time, as we’ve been doing.”

“A slow sort of country!” said the Queen. “Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!”

”

— Lewis Carroll, *Through the Looking-Glass*

**

- [The Ideology Is Not The Movement](#)
- [Against Tulip Subsidies](#)
- [Book Review: Hive Mind](#)
- [SSC Gives A Graduation Speech](#)
- [Basic Income, Not Basic Jobs: Against Hijacking Utopia](#)
- [Bulls**t Jobs \(Part 1 of ∞\)](#)
- [Considerations On Cost Disease](#)
- [What Happened To '90s Environmentalism?](#)
- [The Atomic Bomb Considered As Hungarian High School Science Fair Project](#)
- [Hungarian Education II: Four Nobel Truths](#)
- [Hungarian Education III: Mastering The Core Teachings Of The Budapestians](#)
- [Book Review: Seeing Like A State](#)
- [Book Review: Legal Systems Very Different From Ours](#)
- [Book Review: The Secret Of Our Success](#)
- [Book Review: Secular Cycles](#)
- [Book Review: Ages Of Discord](#)
- [1960: The Year The Singularity Was Cancelled](#)
- [Ars Longa, Vita Brevis](#)

The Ideology Is Not The Movement

Posted on April 4, 2016 by Scott Alexander



I

Why is there such a strong Sunni/Shia divide?

I know the Comparative Religion 101 answer. The early Muslims were debating who was the rightful caliph. Some of them said Abu Bakr, others said Ali, and the dispute has been going on ever since. On the other hand, that was fourteen hundred years ago, both candidates are long dead, and there's no more caliphate. You'd think maybe they'd let the matter rest.

Sure, the two groups have slightly different hadith and schools of jurisprudence, but how many Muslims even *know* which school of jurisprudence they're supposed to be following? It seems like a pretty minor thing to have centuries of animus over.

And so we return again to [Robbers' Cave](#):

The experimental subjects — excuse me, “campers” — were 22 boys between 5th and 6th grade, selected from 22 different schools in Oklahoma City, of stable middle-class Protestant families, doing well in school, median IQ 112. They were as well-adjusted and as similar to each other as the researchers could manage.

The experiment, conducted in the bewildered aftermath of World War II, was meant to investigate the causes—and possible remedies—of intergroup conflict. How would they spark an intergroup conflict to investigate? Well, the 22 boys were divided into two groups of 11 campers, and —

— and that turned out to be quite sufficient.

The researchers’ original plans called for the experiment to be conducted in three stages. In Stage 1, each group of campers would settle in, unaware of the other group’s existence. Toward the end of Stage 1, the groups would gradually be made aware of each other. In Stage 2, a set of contests and prize competitions would set the two groups at odds.

They needn’t have bothered with Stage 2. There was hostility almost from the moment each group became aware of the other group’s existence: They were using our campground, our baseball diamond. On their first meeting, the two groups began hurling insults. They named themselves the Rattlers and the Eagles (they hadn’t needed names when they were the only group on the campground).

When the contests and prizes were announced, in accordance with pre-established experimental procedure, the intergroup rivalry rose to a fever pitch. Good sportsmanship in the contests was evident for the first two days but rapidly disintegrated.

The Eagles stole the Rattlers’ flag and burned it. Rattlers raided the Eagles’ cabin and stole the blue jeans of

the group leader, which they painted orange and carried as a flag the next day, inscribed with the legend "The Last of the Eagles". The Eagles launched a retaliatory raid on the Rattlers, turning over beds, scattering dirt. Then they returned to their cabin where they entrenched and prepared weapons (socks filled with rocks) in case of a return raid. After the Eagles won the last contest planned for Stage 2, the Rattlers raided their cabin and stole the prizes. This developed into a fistfight that the staff had to shut down for fear of injury. The Eagles, retelling the tale among themselves, turned the whole affair into a magnificent victory—they'd chased the Rattlers "over halfway back to their cabin" (they hadn't).

Each group developed a negative stereotype of Them and a contrasting positive stereotype of Us. The Rattlers swore heavily. The Eagles, after winning one game, concluded that the Eagles had won because of their prayers and the Rattlers had lost because they used cuss-words all the time. The Eagles decided to stop using cuss-words themselves. They also concluded that since the Rattlers swore all the time, it would be wiser not to talk to them. The Eagles developed an image of themselves as proper-and-moral; the Rattlers developed an image of themselves as rough-and-tough.

If the researchers had decided that the real difference between the two groups was that the Eagles were adherents of Eagleism, which held cussing as absolutely taboo, and the Rattlers adherents of Rattlerism, which held it a holy duty to cuss five times a day – well, that strikes me as the best equivalent to saying that Sunni and Shia differ over the rightful caliph.

II

Nations, religions, cults, gangs, subcultures, fraternal societies, internet communities, political parties, social movements – these are all really different, but they also have some deep similarities. They're all groups of people. They all combine comradery within the group with a tendency to dislike other groups of the same type. They all tend to have a stated purpose, like electing a candidate or worshipping a deity, but also serve a *very* important role as impromptu social clubs whose members mostly interact with one another instead of outsiders. They all develop an internal culture such that members of the groups often like the same foods, wear the same clothing, play the same sports, and have the same philosophical beliefs as other members of the group – even when there are only tenuous links or no links at all to the stated purpose. They all tend to develop sort of legendary histories, where they celebrate and exaggerate the deeds of the groups' founders and past champions. And they all tend to inspire something like patriotism, where people are proud of their group membership and express that pride through conspicuous use of group symbols, group songs, et cetera. For better or worse, the standard way to refer to this category of thing is "tribe".

Tribalism is potentially present in all groups, but levels differ a lot even in groups of nominally the same type. Modern Belgium seems like an unusually non-tribal nation; Imperial Japan in World War II seems like an unusually tribal one. Neoliberalism and market socialism seem like unusually non-tribal political philosophies; communism and libertarianism seem like unusually tribal ones. Corporations with names like Amalgamated Products Co probably aren't very tribal; charismatic corporations like Apple that become identities for their employees and customers are more so. Cults are maybe the most tribal groups that exist in the modern world, and those Cult Screening Tools make good measures for tribalism as well.

The dangers of tribalism are obvious; for example, fascism is based around dialing a country's tribalism up to eleven, and it ends poorly. If I had written this essay five years ago, it would be titled "Why Tribalism Is Stupid And Needs To Be Destroyed". Since then, I've changed my mind. I've found that I enjoy being in tribes as much as anyone else.

Part of this was resolving a major social fallacy I'd had throughout high school and college, which was that the correct way to make friends was to pick the five most interesting people I knew and try to befriend them. This almost never

worked and I thought it meant I had terrible social skills. Then I looked at what everyone else was doing, and I found that instead of isolated surgical strikes of friendship, they were forming groups. The band people. The mock trial people. The football team people. The Three Popular Girls Who Went Everywhere Together. Once I tried “falling in with” a group, friendship became much easier and self-sustaining precisely because of all of the tribal development that happens when a group of similar people all know each other and have a shared interest. Since then I’ve had good luck finding tribes I like and that accept me – the rationalists being the most obvious example, but even interacting with my coworkers on the same hospital unit at work is better than trying to find and cultivate random people.

Some benefits of tribalism are easy to explain. Tribalism intensifies all positive and prosocial feelings within the tribe. It increases trust within the tribe and allows otherwise-impossible forms of cooperation – remember Haidt on the [Jewish diamond merchants](#) outcompeting their rivals because their mutual Judaism gave them a series of high-trust connections that saved them costly verification procedures? It gives people a support network they can rely on when their luck is bad and they need help. It lets you “be yourself” without worrying that this will be incomprehensible or offensive to somebody who thinks totally differently from you. It creates an instant densely-connected social network of people who mostly get along with one another. It makes people feel like part of something larger than themselves, which makes them happy and can ([provably](#)) improves their physical and mental health.

Others are more complicated. I can just make motions at a feeling that “what I do matters”, in the sense that I will probably never be a Beethoven or a Napoleon who is very important to the history of the world as a whole, but I can do things that are important within the context of a certain group of people. All of this is really good for my happiness and mental health. When people talk about how modern society is “atomized” or “lacks community” or “doesn’t have meaning”, I think they’re talking about a lack of tribalism, which leaves people all alone in the face of a society much too big to understand or affect. The evolutionary psychology angle here is too obvious to even be worth stating.

And others are entirely philosophical. I think some people would say that wanting to have a tribe is like wanting to have a family – part of what it means to be human – and demands to justify either are equally wrong-headed.

Eliezer thinks [every cause wants to be a cult](#). I would phrase this more neutrally as “every cause wants to be a tribe”. I’ve seen a lot of activities go through the following cycle:

1. Let’s get together to do X
2. Let’s get together to do X, and have drinks afterwards
3. Let’s get together to discuss things from an X-informed perspective
4. Let’s get together to discuss the sorts of things that interest people who do X
5. Let’s get together to discuss how the sort of people who do X are much better than the sort of people who do Y.
6. [Dating site](#) for the sort of people who do X
7. Oh god, it was so annoying, she spent the whole date talking about X.
8. X? What X?

This can happen over anything or nothing at all. Despite the artificial nature of the Robbers’ Cove experiment, its groups are easily recognized as tribes. Indeed, the reason this experiment is so interesting is that it shows tribes in their purest form; no veneer of really being about pushing a social change or supporting a caliph, just tribes for tribalism’s sake.

III

Scholars call the process of creating a new tribe “ethnogenesis” – Robbers’ Cave was artificially inducing ethnogenesis to see what would happen. My model of ethnogenesis involves four stages: pre-existing differences, a rallying flag, development, and dissolution.

Pre-existing differences are the raw materials out of which tribes are made. A good tribe combines people who have similar interests and styles of interaction *even before* the ethnogenesis event. Any description of these differences will necessarily involve stereotypes, but a lot of them should be hard to argue. For example, atheists are often pretty similar to one another even before they deconvert from their religion and officially become atheists. They’re usually nerdy, skeptical, rational, not very big on community or togetherness, sarcastic, well-educated. At the risk of going into touchier territory, they’re pretty often white and male. You take a sample of a hundred equally religious churchgoers and pick out the ones who are *most like the sort of people who are atheists* even if all of them are 100% believers. But there’s also something more than that. There are subtle habits of thought, not yet described by any word or sentence, which atheists are more likely to have than other people. It’s part of the reason why atheists *need* atheism as a rallying flag instead of just starting the Skeptical Nerdy Male Club.

The rallying flag is the explicit purpose of the tribe. It’s usually a belief, event, or activity that get people with that specific pre-existing difference together and excited. Often it brings previously latent differences into sharp relief. People meet around the rallying flag, encounter each other, and say “You seem like a kindred soul!” or “I thought I was the only one!” Usually it suggests some course of action, which provides the tribe with a purpose. For atheists, the rallying flag is not believing in God. Somebody says “Hey, I don’t believe in God, if you also don’t believe in God come over here and we’ll hang out together and talk about how much religious people suck.” All the atheists go over by the rallying flag and get very excited about meeting each other. It starts with “Wow, you hate church too?”, moves on to “Really, you also like science fiction?”, and ends up at “Wow, you have the same undefinable habits of thought that I do!”

Development is all of the processes by which the fledgling tribe gains its own culture and history. It’s a turning-inward and strengthening-of-walls, which transforms it from ‘A Group Of People Who Do Not Believe In God And Happen To Be In The Same Place’ to ‘The Atheist Tribe’. For example, atheists have symbols like that ‘A’ inside an atom. They have jokes and mascots like Russell’s Teapot and the Invisible Pink Unicorn. They have their own set of heroes, both mythologized past heroes like Galileo and controversial-but-undeniably-important modern heroes like Richard Dawkins and Daniel Dennett. They have celebrities like P.Z. Myers and Hemant Mehta. They have universally-agreed-upon villains to be booed and hated, like televangelists or the Westboro Baptist Church. They have grievances, like all the times that atheists have been fired or picked on by religious people, and all the laws about pledging allegiance to one nation under God and so on. They have stereotypes about themselves – intelligent, helpful, passionate – and stereotypes about their outgroups – deluded, ignorant, bigoted.

Dissolution is optional. The point of the previous three steps is to build a “wall” between the tribe and the outside, a series of systematic differences that let everybody know which side they’re on. If a tribe was never really that different from the surrounding population, stops caring that much about its rallying flag, and doesn’t develop enough culture, then the wall fails and the members disperse into the surrounding population. The classic example is the assimilation of immigrant groups like Irish-Americans, but history is littered with failed communes, cults, and political movements. Atheism hasn’t quite dissolved yet, but occasionally you see hints of the process. A lot of the comments around “Atheism Plus” centered around this idea of “Okay, talking about how there’s no God all the time has gotten boring, plus nobody interesting believes in God anymore anyway, so let’s become about social justice instead”. The parts of atheism who went along with that message mostly dissolved into the broader social justice community – there are a host of nominally atheist blogs that haven’t talked about anything except social justice in months. Other fragments of the atheist community dissolved into transhumanism, or libertarianism, or any of a number of other things. Although there’s still an atheist community, it no longer seems quite as vibrant and cohesive as it used to be.

We can check this four-stage model by applying it to the Sunni and Shia and seeing if it sticks.

I know very little about early Islam and am relying on sources that might be biased, so don't declare a fatwa against me if I turn out to be wrong, but it looks like from the beginning there were big pre-existing differences between proto-Shia and proto-Sunni. A lot of Ali's earliest supporters were original Muslims who had known Mohammed personally, and a lot of Abu Bakr's earliest supporters were later Muslims high up in the Meccan/Medinan political establishment who'd converted only after it became convenient to do so. It's really easy to imagine cultural, social, and personality differences between these two groups. Probably members in each group already knew one another pretty well, and already had ill feelings towards members of the other, without necessarily being able to draw the group borders clearly or put their exact differences into words. Maybe it was "those goody-goodies who are always going on about how close to Mohammed they were but have no practical governing ability" versus "those sellouts who don't really believe in Islam and just want to keep playing their political games".

Then came the rallying flag: a political disagreement over the succession. One group called themselves "the party of Ali", whose Arabic translation "Shiatu Ali" eventually ended up as just "Shia". The other group won and called itself "the traditional orthodox group", in Arabic "Sunni". Instead of a vague sense of "I wonder whether that guy there is one of those goody-goodies always talking about Mohammed, or whether he's a practical type interested in good governance", people could just ask "Are you for Abu Bakr or Ali?" and later "Are you Sunni or Shia?" Also at some point, I'm not exactly sure how, most of the Sunni ended up in Arabia and most of the Shia ended up in Iraq and Iran, after which I think some pre-existing Iraqi/Iranian vs. Arab cultural differences got absorbed into the Sunni/Shia mix too.

Then came development. Both groups developed elaborate mythologies lionizing their founders. The Sunni got the history of the "rightly-guided caliphs", the Shia exaggerated the first few imams to legendary proportions. They developed grievances against each other; according to Shia history, the Sunnis killed eleven of their twelve leaders, with the twelfth escaping only when God directly plucked him out of the world to serve as a future Messiah. They developed different schools of hadith interpretation and jurisprudence and debated the differences ad nauseum with each other for hundreds of years. A lot of Shia theology is in Farsi; Sunni theology is entirely in Arabic. Sunni clergy usually dress in white; Shia clergy usually dress in black and green. Not all of these were deliberately done in opposition to one another; most were just a consequence of the two camps being walled off from one another and so allowed to develop cultures independently.

Obviously the split hasn't dissolved yet, but it's worth looking at similar splits that have. Catholicism vs. Protestantism is still a going concern in a few places like Ireland, but it's nowhere near the total wars of the 17th century or even the Know-Nothing-Parties of the 19th. Consider that Marco Rubio is Catholic, but nobody [except Salon](#) particularly worries about that or says that it will make him unsuitable to lead a party representing the interests of very evangelical Protestants. Heck, the same party was happy to nominate Mitt Romney, a Mormon, and praise him for his "Christian faith". Part of it is the subsumption of those differences into a larger conflict – most Christians acknowledge Christianity vs. atheism to be a bigger deal than interdenominational disputes these days – and part of it is that everyone of every religion is so influenced by secular American culture that the religions have been reduced to their rallying flags alone rather than being fully developed tribes at this point. American Sunni and Shia seem to be [well on their way to dissolving into each other](#) too.

IV

I want to discuss a couple of issues that I think make more sense once you understand the concept of tribes and rallying flags:

1. Disability

I used to be very confused by disabled people who insist on not wanting a “cure” for their condition. Deaf people and autistic people are the two classic examples, and sure enough we find articles like [Not All Deaf People Want To Be Cured](#) and [They Don't Want An Autism Cure](#). Autistic people can at least argue their minds work differently rather than worse, but being deaf seems to be a straight-out disadvantage: the hearing can do anything the deaf can, and can hear also. A hearing person can become deaf at any time just by wearing earplugs, but a deaf person can't become hearing, at least not without very complicated high-tech surgeries.

When I asked some deaf friends about this, they explained that they had a really close-knit and supportive deaf culture, and that most of their friends, social events, and ways of relating to other people and the world were through this culture. This made sense, but I always wondered: if you were able to hear, couldn't you form some other culture? If worst came to worst and nobody else wanted to talk to you, couldn't you at least have the Ex-Deaf People's Club?

I don't think so. Deafness acts as a rallying flag that connects people, gives them a shared foundation to build culture off of, and walls the group off from other people. If all deaf people magically became able to hear, their culture would eventually drift apart, and they'd be stuck without an ingroup to call their own.

Part of this is reasonable cost-benefit calculation – our society is so vast and atomized, and forming real cohesive tribes is so hard, that they might reasonably expect it would be a lot of trouble to find another group they liked as much as the deaf community. But another part of this seems to be about an urge to cultural self-preservation.

2. Genocide

This term is kind of overused these days. I always thought of it as meaning literally killing every member of a certain group – the Holocaust, for example – but the new usage includes [“cultural genocide”](#). For example, autism rights advocates [sometimes say](#) that anybody who cured autism would be committing genocide – this is of course [soundly mocked](#), but it makes sense if you think of autistic people as a tribe that would be dissolved absent its rallying flag. The tribe would be eliminated – thus “cultural genocide” is a reasonable albeit polemical description.

It seems to me that people have an urge toward cultural self-preservation which is as strong or stronger as the urge to individual self-preservation. Part of this is rational cost-benefit calculation – if someone loses their only tribe and ends up alone in the vast and atomized sea of modern society, it might take years before they can find another tribe and really be at home there. But a lot of it seems to be beyond that, an emotional certainty that losing one's culture and having it replaced with another is not okay, any more than being killed at the same time someone else has a baby is okay. Nor do I think this is necessarily irrational; locating the thing whose survival you care about in the self rather than the community is an assumption, and people can make different assumptions without being obviously wrong.

3. Rationalists

The rationalist community is a group of people (of which I'm a part) who met reading the site Less Wrong and who tend to hang out together online, sometimes hang out together in real life, and tend to befriend each other, work with each other, date each other, and generally move in the same social circles. Some people call it a cult, but that's more a sign of some people having lost vocabulary for anything between “totally atomized individuals” and “outright cult” than any particular cultishness.

But people keep asking me what exactly the rationalist community *is*. Like, what is the thing they believe that makes them rationalists? It can't just be about being rational, because loads of people are interested in that and most of them aren't part of the community. And it can't just be about transhumanism because there are a lot of transhumanists who aren't rationalists, and lots of rationalists who aren't transhumanists. And it can't just be about Bayesianism, because pretty much everyone, rationalist or otherwise, agrees that is a kind of statistics that is useful for some things but not

others. So what, exactly, is it?

This question has always bothered me, but now after thinking about it a lot I finally have a clear answer: rationalism is the belief that Eliezer Yudkowsky is the rightful caliph.

No! Sorry! I think “the rationalist community” is a tribe much like the Sunni or Shia that started off with some pre-existing differences, found a rallying flag, and then developed a culture.

The pre-existing differences range from the obvious to the subtle. A lot of rationalists are mathematicians, programmers, or computer scientists. The average IQ is in the 130s. White men are overrepresented, but so are LGBT and especially transgender people. But there’s more. Nobody likes the Myers-Briggs test, but I continue to find it really interesting that rationalists have some Myers-Briggs types (INTJ/INTP) at ten times the ordinary rate, and other types (ISFJ/ESFP) at only one one-hundredth the ordinary rate. Myers-Briggs doesn’t cleave reality at its joints, but if it measures anything at all about otherwise hard-to-explain differences in thinking styles, the rationalist community heavily selects for those same differences. Sure enough, I am *constantly* running into people who say “This is the only place where I’ve ever found people who think like me” or “I finally feel understood”.

The rallying flag was the Less Wrong Sequences. Eliezer Yudkowsky started a blog (actually, borrowed Robin Hanson’s) about cognitive biases and how to think through them. Whether or not you agreed with him or found him enlightening loaded heavily on those pre-existing differences, so the people who showed up in the comment section got along and started meeting up with each other. “Do you like Eliezer Yudkowsky’s blog?” became a useful proxy for all sorts of things, eventually somebody coined the word “rationalist” to refer to people who did, and then you had a group with nice clear boundaries.

The development is everything else. Obviously a lot of jargon sprung up in the form of terms from the blog itself. The community got heroes like Gwern and Anna Salamon who were notable for being able to approach difficult questions insightfully. It doesn’t have much of an outgroup yet – maybe just bioethicists and evil robots. It has its own foods – MealSquares, that one kind of chocolate everyone in Berkeley started eating around the same time – and [its own games](#). It *definitely* has its own inside jokes. I think its most important aspect, though, is a set of shared mores – everything from “understand the difference between ask and guess culture and don’t get caught up in it” to “cuddling is okay” to “don’t misgender trans people” – and a set of shared philosophical assumptions like utilitarianism and reductionism.

I’m stressing this because I keep hearing people ask “What is the rationalist community?” or “It’s really weird that I seem to be involved in the rationalist community even though I don’t share belief X” as if there’s some sort of necessary-and-sufficient featherless-biped-style ideological criterion for membership. This is why people are saying “Lots of you aren’t even singularitarians, and everyone agrees Bayesian methods are useful in some places and not so useful in others, so what is your community even *about*?” But once again, it’s ~~about Eliezer Yudkowsky being the rightful caliph~~ it’s not necessarily *about* anything.

If you take only one thing from this essay, it’s that communities are best understood not logically but historically. If you want to understand the Shia, don’t reflect upon the true meaning of Ali being the rightful caliph, understand that a dispute involving Ali initiated ethnogenesis, the resulting culture picked up a bunch of features and became useful to various people, and now here we are. If you want to understand the rationalist community, don’t ask exactly how near you have to think the singularity has to be before you qualify for membership, focus on the fact that some stuff Eliezer Yudkowsky wrote led to certain people identifying themselves as “rationalists” and for various reasons I enjoy dinner parties with those people about 10000% more interesting than dinner parties with randomly selected individuals. [nostalgebraist](#) actually summed this up really well: “Maybe the real rationalism was the friends we made along the way.” Maybe that’s the real Shia Islam too, and the real Democratic Party, and so on.

4. Evangelical And Progressive Religion

There seems to be a generational process, sort of like Harold Lee's [theory of immigrant assimilation](#), by which religions dissolve. The first generation believes everything literally. The second generation believes that the religion might not be literally true, but it's an important expression of universal values and they still want to follow the old ways and participate in the church/temple/mosque/mandir community. The third generation is completely secularized.

This was certainly my family's relationship with Judaism. My great-great-grandfather was so Jewish that he left America and returned to Eastern Europe because he was upset at American Jews for not being religious enough. My great-grandfather stayed behind in America but remained a very religious Jew. My grandparents attend synagogue when they can remember, speak a little Yiddish, and identify with the traditions. My parents went to a *really* liberal synagogue where the rabbi didn't believe in God and everyone just agreed they were going through the motions. I got Bar Mitzvahed when I was a kid but haven't been to synagogue in years. My children probably won't even have that much.

So imagine you're an evangelical Christian. All the people you like are also evangelical Christians. Most of your social life happens at church. Most of your good memories involve things like Sunday school and Easter celebrations, and even your bittersweet memories are things like your pastor speaking at your parents' funeral. Most of your hopes and dreams involve marrying someone and having kids and then sharing similarly good times with them. When you try to hang out with people who aren't evangelical Christians, they seem to think really differently than you do, and not at all in a good way. A lot of your happiest intellectual experiences involve geeking out over different Bible verses and the minutiae of different Christian denominations.

Then somebody points out to you that God probably doesn't exist. And even if He does, it's probably in some vague and complicated way, and not the way that means that the Thrice-Reformed Meta-Baptist Church and *only* the Thrice-Reformed Meta-Baptist Church has the correct interpretation of the Bible and everyone else is wrong.

On the one hand, their argument might be convincing. On the other, you are pretty sure that if everyone agreed on this, your culture would be destroyed. Sure, your kids could be Christmas-and-Easter-Christians who still enjoy the cultural aspects and derive personal meaning from the Bible. But you're pretty sure that within a couple of generations your descendants would be exactly as secular as anyone else. Absent the belief that serves as your culture's wall against the outside world, it would dissolve without a trace into the greater homogeneity of Western liberal society. So, do you keep believing a false thing? Or do you give up on everything you love and enjoy and dissolve into a culture that mostly hates and mocks people like you? There's no good choice. This is why it sucks that things like religion and politics are both rallying flags for tribes, and actual things that there may be a correct position on.

5. Religious Literalism

One comment complaint I heard during the height of the Atheist-Theist Online Wars was that atheists were a lot like fundamentalists. Both wanted to interpret the religious texts in the most literal possible way.

Being on the atheist side of these wars, I always wanted to know: well, why wouldn't you? Given that the New Testament clearly says you have to give all your money to the poor, and the Old Testament doesn't say anything about mixing meat and milk, maybe religious Christians should start giving everything to the poor and religious Jews should stop worrying so much about which dishes to use when?

But I think this is the same mistake as treating the Sunni as an organization dedicated to promoting an Abu Bakr caliphate. The holy book is the rallying flag for a religion, but the religion is not itself about the holy book. The rallying flag created a walled-off space where people could undergo the development process and create an independent

culture. That independent culture may diverge significantly from the holy book.

I think that very neurotypical people naturally think in terms of tribes, and the idea that they have to retool their perfectly functional tribe to conform to the exact written text of its holy book or constitution or stated political ideology or something seems silly to them. I think that less neurotypical people – a group including many atheists – think less naturally in terms of tribes and so tend to take claims like “Christianity is about following the Bible” at face value. But Christianity is about being part of the Christian tribe, and although that tribe started around the Bible, maintains its coherence because of the Bible, and is of course naturally influenced by it, if it happens to contradict the Bible in some cases that’s not necessarily surprising or catastrophic.

This is also why I’m not really a fan of debates over whether Islam is really “a religion of peace” or “a religion of violence”, especially if those debates involve mining the Quran for passages that support one’s preferred viewpoint. It’s not just because the Quran is a mess of contradictions with enough interpretive degrees of freedom to prove anything at all. It’s not even because Islam is a host of separate cultures as different from one another as Unitarianism is from the Knights Templar. It’s because the Quran just created the space in which the Islamic culture could evolve, but had only limited impact on that evolution. As well try to predict the warlike or peaceful nature of the United Kingdom by looking at a topographical map of Great Britain.

6. Cultural Appropriation

Thanks to some people who finally explained this to me in a way that made sense. When an item or artform becomes the rallying flag for a tribe, it can threaten the tribe if other people just want to use it as a normal item or artform.

Suppose that rappers start with pre-existing differences from everyone else. Poor, male, non-white minority, lots of experience living in violent places, maybe a certain philosophical outlook towards their condition. Then they get a rallying flag: rap music. They meet one another, like one another. The culture undergoes further development: the lionization of famous rappers, the development of a vocabulary of shared references. They get all of the benefits of being in a tribe like increased trust, social networking, and a sense of pride and identity.

Now suppose some rich white people get into rap. Maybe they get into rap for innocuous reasons: rap is cool, they like the sound of it. Fine. But they don’t share the pre-existing differences, and they can’t be easily assimilated into the tribe. Maybe they develop different conventions, and start saying that instead of being about the struggles of living in severe poverty, rap should be about Founding Fathers. Maybe they start saying the original rappers are bad, and they should stop talking about violence and bitches because that ruins rap’s reputation. Since rich white people tend to be good at gaining power and influence, maybe their opinions are overrepresented at the Annual Rap Awards, and all of a sudden you can’t win a rap award unless your rap is about the Founding Fathers and doesn’t mention violence (except Founding-Father-related duels). All of a sudden if you try to start some kind of impromptu street rap-off, you’re no longer going to find a lot of people like you whom you instantly get along with and can form a high-trust community. You’re going to find half people like that, and half rich white people who strike you as annoying and are always complaining that your raps don’t feature any Founding Fathers at all. The rallying flag fails and the tribe is lost as a cohesive entity.

7. Fake Gamer Girls

A more controversial example of the same. Video gaming isn’t just a fun way to pass the time. It also brings together a group of people with some pre-existing common characteristics: male, nerdy, often abrasive, not very successful, interested in speculation, high-systematizing. It gives them a rallying flag and creates a culture which then develops its own norms, shared reference points, internet memes, webcomics, heroes, shared gripes, even some [unique literature](#). Then other people with very different characteristics and no particular knowledge of the culture start enjoying video

games just because video games are fun. Since the Gamer Tribe has no designated cultural spaces except video games forums and magazines, they view this as an incursion into their cultural spaces and a threat to their existence as a tribe.

Stereotypically this is expressed as them getting angry when girls start playing video games. One can argue that it's unfair to infer tribe membership based on superficial characteristics like gender – in the same way it might be unfair for the Native Americans to assume someone with blonde hair and blue eyes probably doesn't follow the Old Ways – but from the tribe's perspective it's a reasonable first guess.

I've found gamers to get along pretty well with women who share their culture, and poorly with men who don't – but admit that the one often starts from an assumption of foreignness and the other from an assumption of membership. More important, I've found the *idea* of the rejection of the 'fake gamer girl', real or not, raised more as a libel by people who genuinely *do* want to destroy gamer culture, in the sense of cleansing video-game-related spaces of a certain type of person/culture and making them entirely controlled by a different type of person/culture, in much the same way that a rich white person who says any rapper who uses violent lyrics needs to be blacklisted from the rap world has a clear culture-change project going on.

These cultural change projects tend to be framed in terms of which culture has the better values, which I think is a limited perspective. I think America has better values than Pakistan does, but that doesn't mean I want us invading them, let alone razing their culture to the ground and replacing it with our own.

8. Subcultures And Posers

Obligatory David Chapman link. A poser is somebody who uses the rallying flag but doesn't have the pre-existing differences that create tribal membership and so never really fits into the tribe.

9. Nationalism, Patriotism, and Racism

Nationalism and patriotism use national identity as the rallying flag for a strong tribe. In many cases, nationalism becomes ethno-nationalism, which builds tribal identity off of a combination of heritage, language, religion, and culture. It has to be admitted that this can make for some *incredibly* strong tribes. The rallying flag is built into ancestry, and so the walls are near impossible to obliterate. The symbolism and jargon and cultural identity can be instilled from birth onward. Probably the best example of this is the Jews, who combine ethnicity, religion, and language into a bundle deal and have resisted assimilation for millennia.

Sometimes this can devolve into racism. I'm not sure exactly what the difference between ethno-nationalism and racism is, or whether there even *is* a difference, except that "race" is a much more complicated concept than ethnicity and it's probably not a coincidence that it has become most popular in a country like America whose ethnicities are hopelessly confused. The Nazis certainly needed a lot of work to transform concern about the German nation into concern about the Aryan race. But it's fair to say all of this is somewhat related or at least potentially related.

On the other hand, in countries that have non-ethnic notions of heritage, patriotism has an opportunity to substitute for racism. Think about the power of the civil rights message that, whether black or white, we are all Americans.

This is maybe most obvious in sub-national groups. Despite people paying a lot of attention to the supposed racism of Republicans, the rare black Republicans do shockingly well within their party. Both Ben Carson and Herman Cain briefly topped the Republican presidential primary polls during their respective election seasons, and their failures seem to have had much more to do with their own personal qualities than with some sort of generic Republican racism. I see the same with Thomas Sowell, with Hispanic Republicans like Ted Cruz, and Asian Republicans like

Bobby Jindal.

Maybe an even stronger example is the human biodiversity movement, which many people understandably accuse of being entirely about racism. Nevertheless, some of its most leading figures are black – JayMan and Chanda Chisala (who is adjacent to the movement but gets lots of respect within it) – and they seem to get equal treatment and respect to their white counterparts. Their membership in a strong and close-knit tribe screens off everything else about them.

I worry that attempts to undermine nationalism/patriotism in order to fight racism risk backfiring. The weaker the “American” tribe becomes, the more people emphasize their other tribes – which can be either overtly racial or else heavily divided along racial lines (eg political parties). It continues to worry me that people who would never display an American flag on their lawn because “nations are just a club for hating foreigners” now have a campaign sign on their lawn, five bumper stickers on their car, and are identifying more and more strongly with political positions – ie clubs for hating their fellow citizens.

Is there such a thing as conservation of tribalism? Get rid of one tribal identity and people just end up seizing on another? I’m not sure. And anyway, nobody can agree on exactly what the American identity or American tribe is anyway, so any conceivable such identity would probably risk alienating a bunch of people. I guess that makes it a moot point. But I still think that deliberately trying to eradicate patriotism is not as good an idea as is generally believed.

V

I think tribes are interesting and underdiscussed. And in a lot of cases when they are discussed, it’s within preexisting frameworks that tilt the playing field towards recognizing some tribes as fundamentally good, others as fundamentally bad, and ignoring the commonalities between all of them.

But in order to talk about tribes coherently, we need to talk about rallying flags. And that involves admitting that a lot of rallying flags are based on ideologies (which are sometimes wrong), holy books (which are *always* wrong), nationality (which we can’t define), race (which is racist), and works of art (which some people inconveniently want to enjoy just as normal art without any connotations).

My title for this post is also my preferred summary: the ideology is not the movement. Or, more jargonishly – the rallying flag is not the tribe. People are just trying to find a tribe for themselves and keep it intact. This often involves defending an ideology they might not be tempted to defend for any other reason. This doesn’t make them bad, and it *may* not even necessarily mean their tribe deserves to go extinct. I’m reluctant to say for sure whether I think it’s okay to maintain a tribe based on a faulty ideology, but I think it’s at least important to understand that these people are in a crappy situation with no good choices, and they deserve some pity.

Some vital aspects of modern society – freedom of speech, freedom of criticism, access to multiple viewpoints, the existence of entryist tribes with explicit goals of invading and destroying competing tribes as problematic, and the overwhelming pressure to dissolve into the Generic Identity Of Modern Secular Consumerism – make maintaining tribal identities really hard these days. I think some of the most interesting sociological questions revolve around whether there are any ways around the practical and moral difficulties with tribalism, what social phenomena are explicable as the struggle of tribes to maintain themselves in the face of pressure, and whether tribalism continues to be a worthwhile or even a possible project at all.

EDIT: Commenters point out a very similar Melting Asphalt post, [Religion Is Not About Beliefs](#).

Against Tulip Subsidies

Posted on June 6, 2015 by Scott Alexander



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, subsidizing his stupidity?

The pro-subsidy faction declares that the people asking these questions 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 or 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 decupled (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 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.

Book Review: Hive Mind

Posted on December 8, 2015 by Scott Alexander

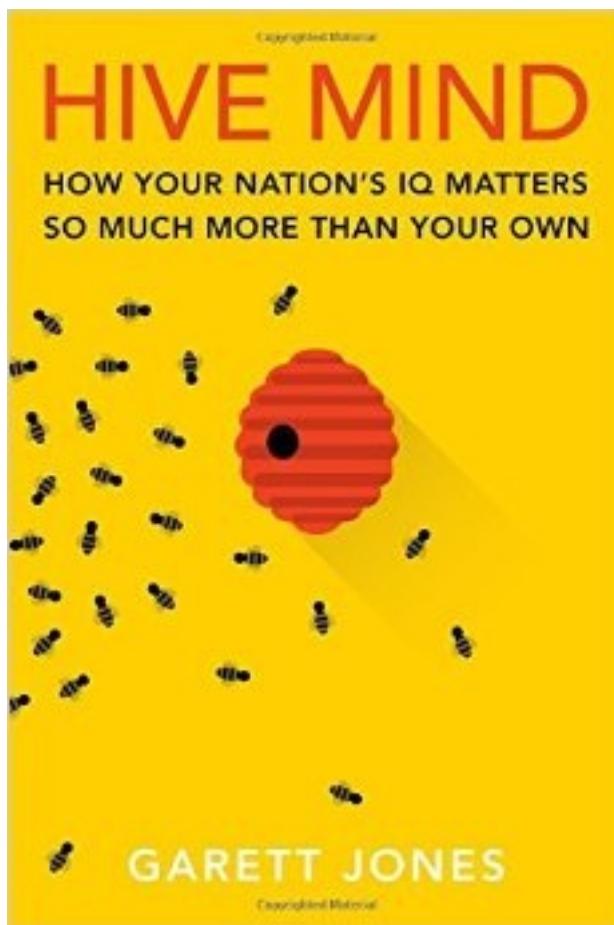


Conflict of interest notice: Author Garrett Jones sometimes reads this blog and is generally great.

☆☆

Garett Jones' book [Hive Mind](#) is classic pop science writing: an intriguing hypothesis, a long parade of interesting studies presented as catchy anecdotes, and not too many follow-up questions.

Its subject (and subtitle) is “why your nation’s IQ matters more than your own”. The gap between rich and poor countries has proven surprisingly resilient, and conventional wisdom is finally getting its head around the idea that something more is going on than a couple of countries getting a head start and the rest of them needing a little time to catch up. Something more than just a temporary lack of capital must be separating the haves from the have-nots, and Jones thinks IQ must be part of the puzzle.



I like my science writing like I like my coffee – COVERED IN BEEES!

He starts with what he calls “the paradox of IQ”. IQ doesn’t matter *that much* on a person-by-person basis. Sure, it’s

correlated with measures of success like personal income, but only weakly. On the other hand, IQ is a very strong predictor of national success – a country's average IQ score correlates very well with whether it's industrialized, rich, First World, and all those nice things. Jones writes:

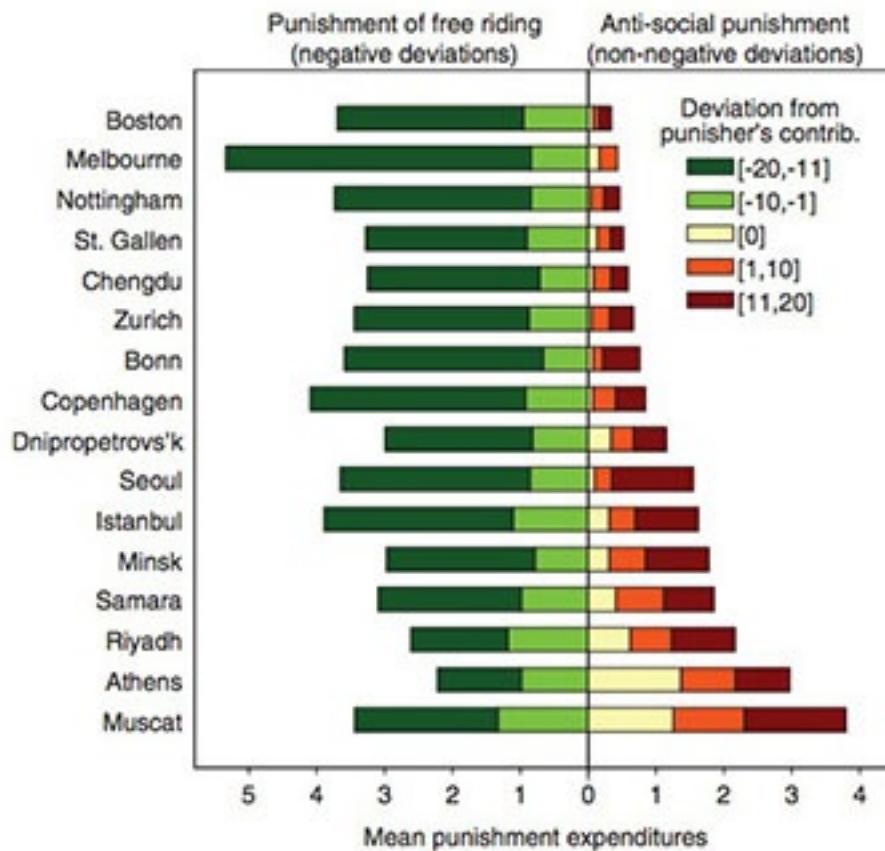
Looking at how individual student test scores predicted those students' wages later in life, they found that individuals with higher test scores earned only slightly more than average within a given country, but nations with higher average test scores grew exceptionally fast. Here again is the paradox of IQ: standardized test scores – whether we call them IQ tests or math tests or something else – predict big national differences but only modest individual differences

I'll talk a little more about that claim in Part II of this review, but for now let's take it seriously and assume causation. Why would IQ matter more for nations than for people?

Jones' theory is that IQ is a measure of people's ability to cooperate in prisoner's dilemma style situations and seek non-zero-sum solutions. Countries where most people have high IQ will come up with mutually beneficial win-win institutions; those where most people have low IQ will be so busy taking advantage of each other and fighting over the pie that they'll never build the institutions necessary for economic growth.

First he reviews research showing that IQ is closely linked to time preference; ie the higher your tested IQ the more likely you are to prefer a big payoff later to a smaller payoff now. For example, in a German experiment a few years ago, participants were offered 100 euros now or X euros in one year; every fifteen IQ points corresponded to a €2.50 change in the value of X necessary for them to accept the latter, even after controlling for education, income, etc. The same thing seems to happen in real life, according to a great study that looked at a natural experiment in the US armed forces. When the military started downsizing after the Cold War, they offered enlisted personnel their choice of various different severance packages – some corresponded to a little money immediately, others to much more money over a longer period. Since the military keeps careful records of the IQ-at-time-of-recruitment of all of its personnel, this was a perfect real-world opportunity to see what happened. The results conformed to theory: IQ predicted tendency to take the longer-term but more lucrative package. There are about twenty studies confirming this result now. And there are also studies showing *national* IQ corresponds with that nation's savings rate, and that individuals who are surrounded by patient frugal people will themselves act more patiently and frugally. If, as the old saying goes, building a good society is about "planting trees in whose shade you will never sit", the people of high IQ nations have a big head start.

Second, he reviews the research from experimental game theory. A series of experiments performed in (of all places) a truck driving school investigated a Window Game. Two players are seated at a desk with a partition between them; there is a small window in the partition. Player A gets \$5 and may pass as much of that as she wants through the window to Player B. Player B may then pass as much as she wants back through the window to Player A, after which the game ends. All money that passes through the window is tripled; eg if Player A passes the entire \$5 through it becomes \$15, and if Player B passes the \$15 back it becomes \$45 – making passing a lucrative strategy but one requiring lots of trust in the other player. I got briefly nerd-sniped trying to figure out the best (morally correct?) strategy here, but getting back to the point: players with high-IQ were more likely to pass money through the window. They were also more likely to reciprocate – ie repay good for good and bad for bad. In a Public Goods Game (each of N players starts with \$10 and can put as much or as little as they like into a pot; afterwards the pot is tripled and redistributed to all players evenly, contributors and noncontributors alike), high-IQ players put more into the pot. They were also more likely to vote for rules penalizing noncontributors. They were also more likely to cooperate and more likely to play closer to traditional tit-for-tat on iterated prisoners' dilemmas. The longer and more complicated the game, the more clearly a pattern emerged: having one high-IQ player was moderately good, but having *all* the players be high-IQ was amazing: they all caught on quickly, cooperated with one another, and built stable systems to enforce that cooperation. In a ten-round series run by Jones himself, games made entirely of high-IQ players had five times as much cooperation as average.



Not technically from the book, but nevertheless fascinating

Third, he reviews the so-called “O-ring theory of teams”, named after the spaceship part that malfunctioned during the *Challenger* explosion. The theory is: suppose that a spaceship requires a million different parts to work. This is more than just a million times harder than building a spaceship that requires one part to work. If you have a spaceship engineer who can build a part and be 99% sure she’s gotten it right, this is probably good enough for the one-part spaceship: a 99% success rate for a spaceship sounds pretty good. But if the spaceship uses a million parts and they all have to be perfect, your chances of success with a million such engineers is $0.99^{1000000}$, aka zero. You had better find some better spaceship engineers! This gives high-IQ societies a big leg up when they’re working on complicated projects; a low-IQ society may have some high-IQ individuals who can do good work on their own, but including even a single low-IQ individual on a spaceship will screw it up big-time. This theory implies that people will end up segregated by ability. Imagine you have four spaceship engineers, two of whom are good (99% accuracy) and two of whom are mediocre (50% accuracy), and you want to build two two-part spaceships. If you pair up one good and one mediocre engineer on each, each of your spaceships will have a $0.99 * 0.50 = 49\%$ chance of success, for a total of 0.98 projected successful spaceships. If you have the two good engineers work together on one ship, and the two mediocre ones work together on the other, you’ll have a 98% success rate on the first one and a 25% success rate on the second one, for a total of 1.23 projected successful spaceships. You’ve gained a quarter-spaceship just by segregating your engineers by ability. The more high-IQ people you have, the easier this is and the more you can devote your economy to complex things like million-part spaceships. The more low-IQ people you have, the harder this gets and the more your economy sticks to high-failure-tolerance but less lucrative products.

Finally, high-IQ people are smart (citation needed). They tend to know what policies are good and what policies are bad and vote for the good ones. Here Jones cites Bryan Caplan’s [*The Myth of the Rational Voter*](#) a lot, showing that voters aren’t very good at figuring out their own self-interest.

But he has a more positive spin: high-IQ voters *do* seem good at this. As a GMU economist, Jones' measure for "are people voting rationally" is of course "how pro-free-market are they?", and he finds that high IQ predicts pro-market attitudes pretty strongly and in fact better than years-of-education. In controlled experiments higher-IQ people were more likely to be able to admit that a test article contradicted their political bias, and in some countries (although not the US) high-IQ people are more likely to vote.

Then he ties all of this together into a kind of [stationary-bandit framework](#), where government starts with selfish warlords who want to exploit the populace.

"They say all government started w/ stationery bandits."

"Really?"

"Yeah. They had to steal enough nice paper to write a constitution on."

— [Scott Alexander \(@slatestarcodex\)](#), December 8, 2015

If you're a high-IQ selfish warlord, and your oppressive ministers are likewise high-IQ, you have enough patience to realize that if you leave the capitalists alone to do their thing instead of confiscating their wealth immediately, in a couple of years they'll have even *more* wealth you can confiscate. And if some kind of conflict comes up and threatens to lead to civil war, you are good at negotiating win-win solutions where everyone cooperates to increase the size of the pot. Jones lists a bunch of political situations that map to iterated prisoner's dilemmas – for example, do both parties respect election results, or does the loser try to start a fight over it every time they're forced out of government? Do bureaucracies try to run the country efficiently, or do they jockey for power against each other? Do military branches work together during operations, or does each one try to seize glory for its own leaders? If you have a high-IQ country, these problems have a way of just solving themselves – and sure enough, IQ scores correlate nicely with the Corruption Perceptions Index. And businesspeople know this, so they are happy to start complicated long-term projects in the countries with a history of tolerating such projects and not killing the golden-egg-laying geese.

Jones doesn't go too deep into policy prescriptions, but he does mention two consequences of his theory. First, he's a big fan of the Flynn Effect (secular trend of rising IQs) and thinks that countries ought to encourage this so that their national IQ gets higher and they can have more effective institutions – unfortunately, he doesn't know what's causing the Flynn Effect any more than anyone else does, so this sort of reads as "keep doing the thing we don't know how we're doing". He does think that eliminating lead will help (did you know sub-Saharan Africa was the last region to ban leaded gasoline, all the way in 2006?) and he has the usual hopes for nutritional, educational and health interventions.

But of course the part everyone's talking about is immigration. This is *not* a major focus of the book. Jones actually spends more time talking about all the benefits of immigration than anything else:

About a decade ago, dozens of American economists signed an open letter in support of more immigration. The letter touched on many points: that less-skilled immigrants appear to push down the wages of US born citizens little if at all, that immigration helps rich country economies in ways that don't show up in official statistics, and that the biggest beneficiaries of less-skilled immigration are the immigrants themselves, whose lives are often transformed from a nightmares of dollar-a-day poverty to a realm of modest comfort, health and safety. The diplomatically crafted letter, circulated by the Independent Institute, was signed by economists on the left and the right. I've always been glad I signed this letter: it sums up the great promise of immigration... for people who care about ending the deepest poverty, migration should be at the top of the list of potential cures.

But he does devote about one-and-three-quarters pages to his concerns:

The economics of less-skilled immigration to richer, more productive countries are reasonably clear: life-changing good news for the immigrant with only fairly small effects one way or the other on so-called “native” less-skilled workers. That’s true when we look at the short run or when we look across towns and cities within the same country. And crucially, these studies hold politics aside and assume that less-skilled immigrants don’t have an effect on a high-skill nation’s government institutions. But if there’s something we’ve seen in previous chapters, or something we’ve seen in Bryan Caplan’s work on the link between voter education and voter beliefs, if there’s something we’ve seen in the cross-country studies that find that higher national average test scores tend to predict lower average levels of corruption and in the philosophical debates over epistocracy, it’s that good politics appears to depend on reasonably well-informed citizens. With this we come to a central tension of immigration among the currently less-skilled: the possible – I emphasize possible – effect on long-run institutions. Will less-skilled immigrants tend to vote for policies that will weaken the wealth-creating opportunities they’ve enjoyed? Or will less-skilled immigrants and their descendants instead build up high levels of human capital, perhaps raising the average information level of voters?

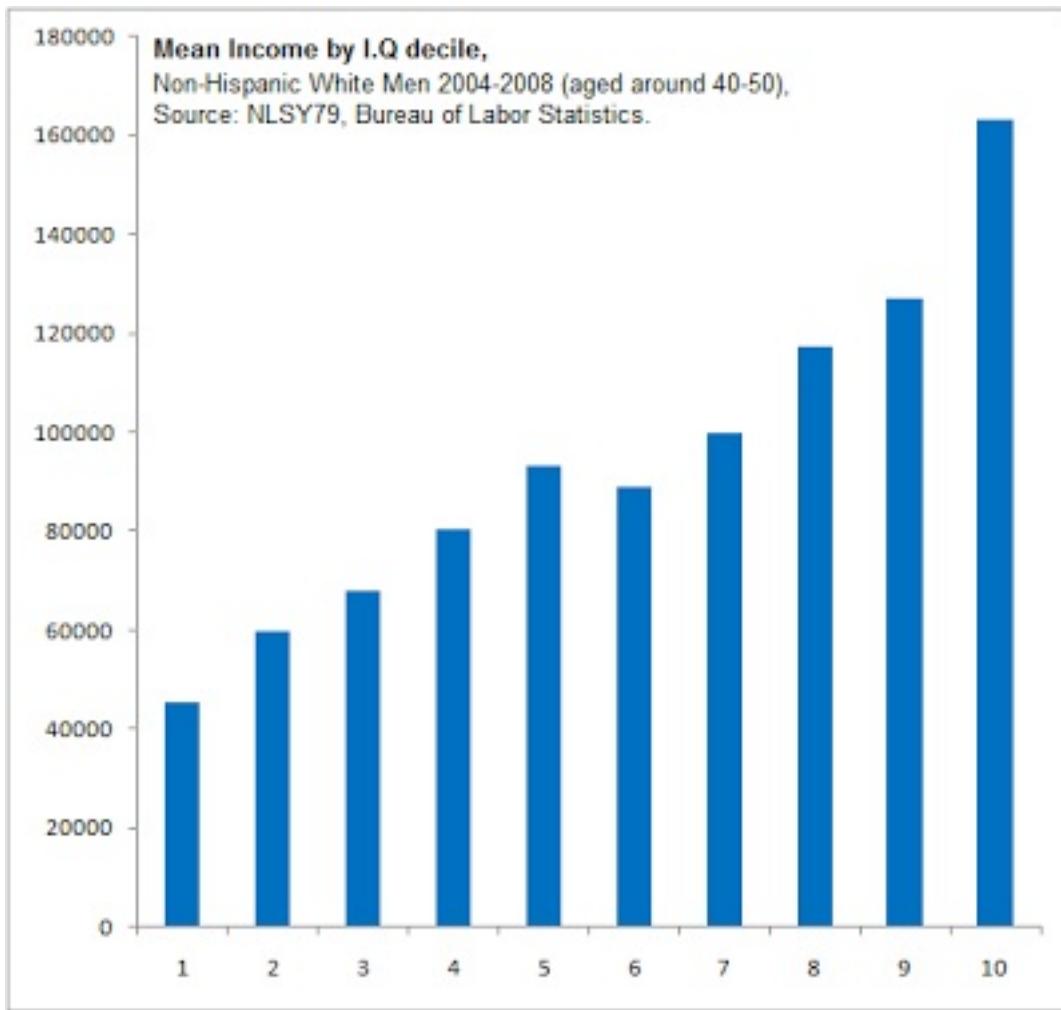
The whole paragraph has the feeling of somebody being dragged over a bed of hot coals, from the insistence on referring to unskilled immigrants as “currently less-skilled” and natives as “so-called native less-skilled workers” to the odd proposal at the end that maybe for some reason less-skilled workers will actually *raise* the average information level of voters, because *who really knows?* This book is emphatically not *The Bell Curve*. It’s a book about science which is deeply annoyed that it might have controversial political implications and tries to avoid them as carefully as possible, generally successfully.

II

There were some parts of this book that I did not find convincing, or that at least left me with further questions.

First, *Hive Mind*’s “central paradox” is why IQ has very little predictive power among individuals, but very high predictive power among nations. Jones’ answer is [long complicated theory of social cooperation]. Why not just “signal-to-noise ratio gets higher as sample size increases”?

Jones’ paradox was very similar to the question I asked in [Beware Summary Statistics](#), except I was wondering not about nations, but about abstracted IQ deciles:



On a personal level, IQ has modest predictive power. But if you average out thousands of IQ 90 people, thousands of IQ 100 people, and thousands of IQ 110 people, the IQ-income relationship will become very clear. At this level of abstraction, it is no longer fair to describe it as “modest”.

That first block corresponds to people of about IQ 80, the last block to people of about IQ 120. As you move from 80 to 120, income practically quadruples. And this is within the United States, where we've got all sorts of minimum wage laws and so on likely to dampen the effects.

Or to give a more natural example – Jews have 10-15 points higher IQ than WASPs in America, and make about twice as much money. This happens even though most Jews do not solely interact with other Jews or make their own institutions – there are few opportunities for them to form a hive mind. Their individual IQ differences, once aggregated, seem to produce the strong effect.

There is much-larger between-country variance in income than between-individuals-in-country variance in income, but it doesn't seem obvious to me that the percent of between-country variance explained by national IQ is larger than the percent of between-individuals-in-country explained by personal IQ once factors like personal job choice (I could have been an investment banker, but I would rather be an artist) that countries don't have to deal with is abstracted out. If the amount-variance-explained between nations and between individuals were equal after adjusting for that factor, there would not be any need to posit hive mind-type effects.

EDIT: Above heavily edited for clarity and correctness after originally being much weaker argument in same direction.
See [here](#). Some complicated discussion of this going on [here](#), see especially Pseudoerasmus' comments

Although it may be that there's a national effect stronger than the aggregated-individual effect, I feel like this is something Jones should have had to prove, rather than relying on a "look, it's obvious!" based on unaggregated-individual numbers.

Second, fine. Let's assume he proves to our satisfaction that the national IQ-income correlation is sufficiently stronger than individual ones to require explanation. Now we get to my biggest gripe with this entire book. *How do we know the direction is IQ → development rather than development → IQ?!* Jones lays out exactly the set of assumptions that make reverse causation most plausible. He dedicates an entire chapter to the Flynn Effect, how he thinks it's real, how he thinks it's a big deal, without mentioning whether the gains might not be on g. Time and time again, Jones hammers how countries' IQs increase as they develop further. Then he shows us a graph of IQ-development correlation and just assumes the causation is bidirectional. Well, why not just development → IQ?

This isn't just about me. I suspect Jones is right – though I'm not entirely sure of it – and sufficiently biased in favor of that position to be happy to follow it and see where it leads. I'm asking for anybody who reads this book without already being interested in IQ. *Hive Mind* is clearly pitched at a smart layperson audience, and any smart layperson who reads this book ought to have exactly that question, asked with exactly that many capital letters and explanation points. Any reader who doesn't immediately stand on a chair and shout "Where is the evidence against reverse causation?" is not a reader that Garrett Jones should want. But any reader who does that will not find an answer.

I'm just sayin', everyone that confuses correlation with causation eventually ends up dead.

— [Siberian Fox \(@SilverVVulpes\)](#), September 14, 2015

All I can say in his defense is that a good defense against this accusation would probably have to get very deep into the causes of IQ, exactly the subject Jones is carefully trying to avoid. I understand his reluctance to approach this subject and respect his strategic decision. All I can say is that it leaves a hole in his argument big enough to sail an oil tanker through.

EDIT: Jones responds [here](#)

Third, and this isn't such a problem as the others but it left me curious – how do we go from the short, few-player games that make up most of the book's experiments, to the multi-generation million-player games that make up real countries?

I have two concerns here. First, Jones says that:

The one study of which I'm aware that finds that higher-IQ individuals are more cruel and less cooperative is a study of a one-shot prisoners' dilemma, something much like the true criminal's prisoner's dilemma... this is the only setting I know of in which high scorers are more brutal than low scorers... in a one-shot environment, if it's either steal or be robbed, and if the players will never see each other again, then I'd expect higher-IQ individuals to figure out what setting they're in and act shrewdly, not cruelly.

He returns to this theme a few times. High-IQ people don't cooperate because they're nicer (which, indeed, personality tests for niceness do not show). They cooperate because they're smarter and so they know cooperation really is a better and more win-win way to do things.

This is 100% true in an iterated prisoners' dilemma, but not necessarily true in a country. Suppose you're a president with a four year term. You can either pillage the country as best you can and take whatever bribes you can get, or invest in genuinely building a better country for your descendants. Assuming you are merely the sort of shrewd

cooperator who cooperates on iterated prisoners' dilemmas but defects on one-shots, you'll pillage the country – it probably has term limits and you only need to pillage once to get very rich.

Likewise, suppose you're a mid-level bureaucrat in Washington, of the type that there are tens of thousands of. If you behave dishonorably, you can amass a small empire and make some money. If you behave honorably, then maybe America does very well as a country down the line, but that effect is aggregated over thousands of bureaucrats, so it's not like you're *really* growing the pot that much. Once again, if you are merely shrewd and not genuinely altruistic, you'll defect.

Jones tries to take the easy way out on the deriving-ethical-behavior thing here, saying that ethical behavior *really is* the most self-serving option in the long-term, and all you need is people smart enough to realize it. To that I can only say: no it isn't. In a game of two or three people where everyone sees everyone else's results, your contribution may grow the pot enough to be worth the short-term losses from not defecting. In a game of thousands of bureaucrats or millions of citizens, not so much. There are ideas like TDT and superrationality that try to bridge this gap, but I think Jones tries to cross it without those ideas and is left floundering.

One more thing on this topic: maybe it was in the original studies and I just didn't look deep enough, but I wonder how much of this is just understanding the game. The Iterated Prisoner's Dilemma is kind of complicated, and if you're stupid you may not be able to grasp the logic behind why cooperation is sometimes the better option. If you explained everything very carefully to all participants, had them play a couple of games both ways so they got a feel for it, and had a Professor of Economics give a lecture on why cooperation was probably the best option, would high-IQ people still succeed more because of some innate cooperative tendency? Or would everyone else have figured out their secret and robbed them of their advantage?

People usually have a pretty good grasp of things that are going on in society. Jones compares marriage to a prisoner's dilemma (where the optimal C-D outcome is "you cheat but your spouse stays faithful"). But people understand the terms of marriage, cultural evolution has had a long time to come up with beliefs and mores about marriage that even people too stupid to come up with them on their own follow, and some kind of complicated new game may not be the best analogue to the marriage problem.

III

Jones ends the book with the following observation:

The best guess is that the cognitive skill of elites really does matter more than the nation's average score. When it comes to institutional quality, Potrafke and I found that the cognitive skills of the top 5 percent did the better job of predicting property-rights friendly institutions, although the nation's average score also did a reasonably good job as a predictor... for the time being it's reasonable to start with the belief that a nation's top performers matter more for the economy than a nation's average performers.

Well, *that's* interesting. All of this stuff about immigration and on how maybe we shouldn't have open borders, and it turns out that as long as the top five percent are smart, everything is okay.

I would really like to see more on this. If America has higher IQ than Britain, but the members of Parliament have higher IQ than the members of Congress, which country will do better? What about a colonial nation where the administrators are from a nation that has a completely different IQ than the population? What about countries that have multiple mostly-segregated populations with different IQs? How much does the IQ of the government matter versus the IQ of the population itself?

(and now I wonder if Jones has read La Griffe on smart fractions [[1](#),[2](#)])

Come to think of it, doesn't every nation have some pretty smart people at the highest echelons? Sub-Saharan Africa may be in the IQ doldrums, but we all know African economists, statesmen, etc whose work is top-notch. Doesn't Jones' call to raise national IQs with the Flynn Effect seem less pressing now? Haven't the elites of Third World countries already probably been Flynn-ified, since they usually get good food, good medical care, and good education? Should we worry the Flynn Effect won't help those countries further? Or should we hope that if we merely raise the IQ of a few people, that will be enough and we won't have to have a mass nationwide campaign? (calling all CRISPR enthusiasts...)

Overall, I thought this book showcased some really neat results, had some good economics in it, and was very readable, but I didn't come out of it feeling like its thesis was very proven.

SSC Gives A Graduation Speech

Posted on May 23, 2014 by Scott Alexander



Trigger warning: 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 eucore 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*.

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!

Basic Income, Not Basic Jobs: Against Hijacking Utopia

Posted on May 16, 2018 by Scott Alexander



*
*
*

Some Democrats angling for the 2020 presidential nomination have a big idea: [a basic jobs guarantee](#), where the government promises a job to anybody who wants one. Cory Booker, Kirsten Gillibrand, Elizabeth Warren, and Bernie Sanders are all said to be considering the plan.

I've pushed for a basic income guarantee before, and basic job guarantees sure sound similar. Some thinkers have even compared the two plans, pointing out various advantages of basic jobs: it feels "fairer" to make people work for their money, maybe there's a psychological boost from being productive, you can use the labor to do useful projects. Simon Sarris has [a long and excellent article](#) on "why basic jobs might fare better than UBI [universal basic income]", saying that:

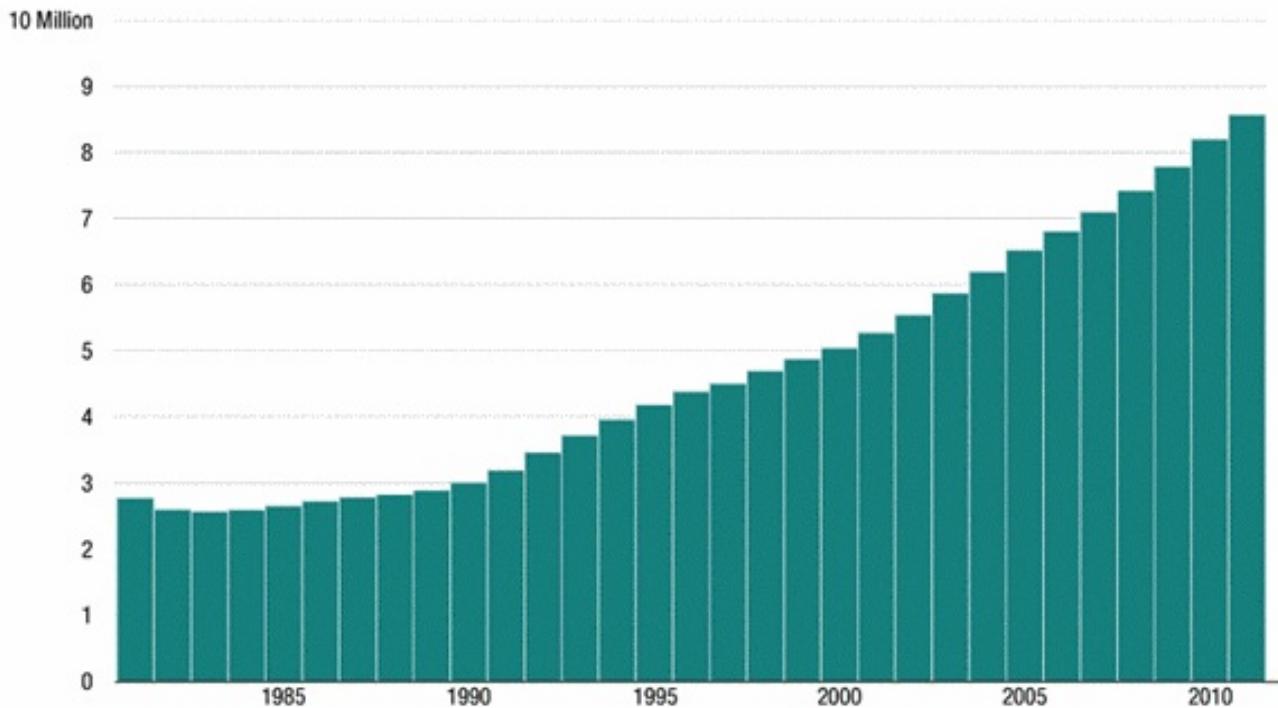
UBI's blanket-of-money approach optimizes for a certain kind of poverty, but it may create more in the long run. Basic Jobs introduce work and opportunity for communities, which may be a better welfare optimization strategy, and we could do it while keeping a targeted approach to aiding the poorest.

I am totally against this. Maybe basic jobs are better than nothing, but I have an absolute 100% revulsion at the idea of implementing basic jobs as an alternative to basic income. Before getting into the revulsion itself, I want to bring up some more practical objections:

1. Basic jobs don't help the disabled

Disability has doubled over the past twenty years and continues to increase.

Number Of Former Workers On Disability



Experts disagree on how much of the rise in disability reflects deteriorating national health vs. people finding a way to opt out of an increasingly dysfunctional labor market, but everyone expects the trend to continue. Any program aimed at the non-working poor which focuses on the traditionally unemployed but ignores the disabled is only dealing with the tip of the iceberg.

The current disability system has at least three major problems which I would expect basic income to solve.

First, the disability application process is a mess. Imagine the worst DMV appointment you've ever had to obtain the registration to a sketchy old car you got from a friend, then multiply it by a thousand – then imagine you have to do it all while being too disabled to work. Even clear-cut applications can take months to go through, inflicting an immense burden on people who don't know where their money is coming from during that time. And people with harder-to-prove conditions like mental illness and chronic pain might require multiple appeals – dragging the process out for years – or never get it at all. The disabled people I have talked to generally hate everything about this.

Second, disability is becoming a catch-all for people who can't find employment. This is a useful function that needs to be served. But right now, it involves unemployed people faking and exaggerating disability. This rewards liars and punishes the honest. If society labels the system "FOR DISABLED PEOPLE ONLY", basic fairness – to the disabled, to taxpayers, and to honest workers who aren't gaming the system – require them to gatekeep entry. Right now they spend lots of time and money on gatekeeping and still mostly fail. But any attempt to crack down would exacerbate the first problem, the one where real disabled people have to spend months or years in a Kafka novel before getting recognized.

Third, because of the first and second problems disabled people feel like they constantly have to prove themselves. Sometimes they'll have good days – lots of conditions are relapsing-remitting – and they'll want to go play in the park or something. Then they have to worry that some neighbor is going to think "well, that guy looks pretty healthy", take a photo, and they'll end up as one of those stories with headlines like SO-CALLED DISABLED PERSON CAUGHT

PLAYING SPORTS IN PARK. Other times it's a bureaucratic issue. I had a patient who, after a few years on disability, recovered enough that he thought he could work about ten hours a week. When he tried to make it happen, he learned he would lose his disability payments – apparently if you can work at all the government doesn't believe you're really disabled – and ten hours a week wasn't enough to support himself. So he cancelled the new job and didn't work at all.

As long as you have a system whose goal is to separate the “truly” disabled people from the fakers, you’re going to run into problems like these. But refuse to gatekeep, and you have an unjust system where anyone who wants to lie can get out of work while their more honest coworkers are left slaving away all day. Basic income cuts the Gordian knot by proposing that everyone is legally entitled to support, whether they’re disabled or not. Disabled people can get their money without gatekeeping, and there’s no reward for foul play.

Basic jobs abandons this solution and takes us right back to the current system. If you’re abled enough to perform a government job, you’ve got to do it. Who decides if you’re abled enough? The Kafkaesque gatekeepers. And so we get the same bureaucratic despair, the same attempts to cheat the system, and the same perverse incentives.

And the number of disability claims keeps rising. Remember, a lot of economists think that the flight away from work and toward disability comes from people voting with their feet against exactly the kind of low-paying unpleasant jobs that basic jobs advocates want to offer everybody. Expect them to vote against those too, with no clear solutions within the basic jobs paradigm.

2. Basic jobs don't help caretakers

And another 10% to 15% of the jobless are people caring for their sick family members.

This is unavoidable and currently uncompensated. The AgingCare Caregiver forum says their "[number one question](#)" is whether people who need to take time off work to care for a sick or elderly parent can get money. The only answer they can provide is "if the person you're caring for has money or insurance, maybe they can pay you". If they don't, you're out of luck. [EDIT: apparently [some states](#) do offer some money for this].

Right now our society just drops the ball on this problem. I don't blame it; giving people money to care for family members would be prohibitively expensive. It would also require a gatekeeping bureaucracy that would put the disability gatekeeping bureaucracy to shame. Not only do they have to assess if someone's really unable to subsist without care, they also have to decide who gets to take the option for which relatives. I have a second cousin some number of times removed who's very disabled; can I quit my job and get paid a reasonable salary to take care of him? What if I tell you I've never met him or even talked to him on the phone, and just have my grandmother's word for it that he exists and is sick? What exactly counts as caretaking? If I go visit my second cousin once a day for an hour to make sure he hasn't gotten any sicker than usual, should the government pay me a full salary? What if actually doing that is 100% vital to my second cousin's continued survival and I wouldn't be able to do it consistently while holding down a job? You are never going to be able to make a bureaucracy that can address all these issues fairly.

Basic income cuts the knot again, giving everyone enough money that they can take care of sick or aging friends or relatives if they so choose. You don't have to justify your choice to provide this level of care (but not that level) to the government. You can just do what needs to be done.

Basic jobs once again drops the ball on this problem. If your mother is dying, you can't be there to help her, because the government is going to make you dig ditches and fill them in again all day to satisfy people's worry that somebody somewhere might be getting money without doing enough make-work to “deserve” it.

3. Basic jobs don't help parents

Everything above, except this time you're a single parent (or a double parent whose spouse also works) and you want to take care of your child. If you could afford daycare, you probably wouldn't be the sort of person who needs to apply for a guaranteed basic job. What do you do?

I know what the basic jobs people's solution to this is going to be: free daycare for all! Okay. So in addition to proposing the most expensive government program ever invented, you want to supplement it by passing the second most expensive government program ever invented, at the same time? Good luck.

But even aside from this, I want us to step back and think about what we're doing. I have met people – mostly mothers, but some fathers too – who are heartbroken at the thought of missing the best years of their children's lives grinding away at a 9 to 5 job, stuck in traffic commuting to their job, or being too tired to spend time with them after they get home from their job. These people miss their kids' first steps, outsource watching their first words to underpaid daycare employees, and have to choose between attending their kids' school plays and putting food on the table.

And if we check the Treasury and decide that we, as a society, don't have enough money to solve this problem – then whatever, we don't have enough money to solve this problem.

But I worry we're going to check and find we have more than enough money. But somebody is going to be so excited about making poor people do busy-work to justify their existence, that we're going to insist on perpetuating the problem anyway. And if that forces us to pay for universal free daycare, we're *going to be spending extra money just to make sure we can perpetuate the problem as effectively as possible*. We're going to be saying "We could give basic income for \$800 billion, or basic jobs plus universal daycare for \$900 billion. And that extra \$100 billion? That's the money we spend to make sure you're digging ditches and filling them in all day, instead of getting to be at home spending time with your kids."

4. Jobs are actually a big cause of poverty

Poor people's two largest expenses are [housing and transportation](#).

Guaranteed jobs have to be somewhere. Most of them will be in big cities, because that's where everybody is. The ones in the country will be few and far between.

That means to get to your government-mandated job, you'll either need to live in the big city or have a car. Living in the big city means tripling your monthly rent. Having a car means car payments, insurance payments, repair payments, gas payments, and incidentals.

When I first started working with poor patients, I was shocked how many of the problems in their lives were car-related. For well-off people like me, having a car is background noise; you buy or lease it for a reasonable price, then never worry about it again. Poor people can't afford to buy and don't always have good enough credit to lease. They tend to get older, sketchier cars that constantly break down. A constant complaint I heard: "My car broke, I can't afford repairs, and I'm going to get fired if I can't make it to my job". Some of them can't afford insurance and take their chances without it. Others have had various incidents with the police that cost them their license, but they can't just not show up to work, so they drive anyway and hope they don't get arrested.

Then there are the little things. Your work doesn't have a break room, so you've got to eat out for lunch, and there goes a big part of your food budget. Your work demands a whole new set of business clothes, so there's double your clothing budget. You can't attend things during normal business hours, so you have to pay extra for out-of-hours services.

And then there's all of the problems above. You can't take care of your children anymore, so you've got to pay for

daycare or a nanny or an Uber to take them to their grandparents' house. You can't take care of your sick parents anymore, so you've got to pay for a home health aide to come in and look after them. You get job-related strain or stress, and there's the cost of a doctor's appointment.

And then there are the fuzzier things. If you've just spent the entire day at work, and you're really exhausted, and you never get any time to yourself, maybe you don't have the energy left to drive to the cheaper supermarket on the other end of town. Maybe you don't have the time to search for the absolute best deal on the new computer you're getting. Maybe you don't have the willpower to resist splurging and giving yourself one nice thing in your life of wage slavery. All of this sounds kind of shameful, but they're all things that my patients have told me and *things that I do myself sometimes despite my perfectly nice well-paying job*.

5. Basic jobs may not pay for themselves by doing useful work

I once read an economist discussing why unemployment exists at all. That is, there are always people who would like to have someone clean their house, take care of their children, or come to their house and cook them food. And there are always businesses that would like their floors a little cleaner, or their customers served a little faster, or one more security guard to keep everything safe. Surely they would pay *some* amount of money to get these jobs done? And surely *some* homeless person would rather take that small amount than starve on the streets? So why are there still unemployed people?

One answer must be the minimum wage, but how come this happens even in times and places where minimum wages are absent or easy to evade?

The economist suggested that not all employees are net positive. Employees can steal from you, offend your customers, or be generally weird and smelly and ruin the atmosphere. They can be late or not show up at all – and if you made plans depending on their presence, that can be worse than your never hiring them in the first place. A bad nanny can traumatize your kid. A bad maid can break your priceless vase. A litigious employee can take you to court on false charges. Somebody who's loud and curses at you and constantly smells of marijuana can just make you a little more stressed and unhappy all the time.

So if you have a job that only produces 1 utility, but a bad employee in that job will cost you 10 utility, and there's a 10% chance any employee you get will be bad – then you're not going to fill that job no matter how low a salary people are willing to work for.

How bad can employees get? Please read these AskReddit links. They're slightly off-topic, but they're going to give you information you can't get any other way:

- [AskReddit: Bosses of Reddit, what was your worst employee like?](#)
- [Managers of Reddit, who was your worst employee?](#)
- [What is the worst employee you have had to put up with?](#)
- [Who's the worst coworker you ever had?](#) It's safe to say they can get pretty bad.

I know many unemployed people who are amazing virtuous hard-working folks. But I also know the unemployed guy who lives in a cardboard box by the BART station, is surrounded by a protective shell of discarded beer cans, and shouts "GRAAAAGH" at passers-by for inscrutable reasons. And the amazing virtuous hard-working folks have a decent shot at getting a job in the private sector eventually, but the guy who shouts "GRAAAAGH" never will. Your population of basic-job-needy is going to be disproportionately composed of people who don't fit into the regular workforce. How do you think that will turn out?

I worry some people think choosing basic jobs over basic income means free labor. Like, if you were going to pay someone a basic income of \$10K/year, but the market value of their labor is \$8K/year, you could employ them running a soup kitchen, get that \$8K of value, and then you're really only "losing" \$2K/year.

I am less sanguine. If you pay people \$10K/year, you're only losing \$10K/year. If you employ them to run a soup kitchen, and the soup kitchen has to keep closing because of hygiene violations, or gets hit with a sexual harassment lawsuit because someone groped a customer, or burns down because someone left the stove on, or loses all its customers because the manager shouts "GRAAAAGH" at everybody who asks for soup – then you're losing more.

6. Private industry deals with bad workers by firing them; nobody has a good plan for how basic jobs would replace this

Suppose someone *does* accidentally leave a stove on and burn down the soup kitchen. You transfer them to an agricultural commune and they crash the tractor into a tree. You transfer them to some kind of low-risk paper-pushing job, but they're late to work every day and skip it entirely once or twice a week, and important papers end up tragically un-pushed. After a while, you decide they are too incompetent to add non-negative value to any of the programs on offer. What do you do with them?

If you fire them, then you're not a basic jobs guarantee. You're a basic-jobs-for-skilled-workers-whom-bosses-like guarantee. We already have one of those – it's called capitalism, maybe you've heard of it. But a real solution to poverty would have to encompass everybody, not just people who are good at working within the system.

And if you don't fire them, what's your plan? Accept a certain level of burning-things-down, customer complaints, coworker complaints, and unexcused absences? Let them make everybody around them miserable? Turn your soup kitchen into some kind of federal disaster area because you're absolutely committed to letting every single human being in the United States work there?

Or transfer them to a job in a padded room putting blocks in stacks and knocking them down again, in a way that inconveniences nobody because nobody cares about it? Abandon all pretense at creating anything other than busy-work for poor people out of an all-consuming desire to make sure nobody can live comfortably unless they have spent forty hours of every week in boredom and misery?

Or offer *these* people a basic income, and let all your other employees hate you for giving incompetent people leisure time at home with their family while the hard workers dig ditches all day?

This isn't speculation about some vague future. These questions get played out all around the country in our existing "government must take everyone no matter how little they want to be there" institution, ie public school. Here's a quote from a reader the last time we discussed the public school system.

I was friends with a guy who briefly worked as a teacher at a public high school in central DC (I'm 80% sure it was Cardozo High). He had an education background thanks to spending several years working as a youth camp counselor and as an after-school program counselor, and that was sufficient to qualify him for DCPS' abbreviated teacher training program (such a thing existed in 2009 when he did it; I'm unsure if it is still around). During the training program, I remember him speaking about his enthusiasm for the teaching skills he was learning and about his eagerness to put them to use (in retrospect, I think some of this was a nervous attempt to convince himself the job wouldn't be bad). After a break of several months, we spoke again, and he was almost totally disillusioned with the job and was already thinking of quitting. This is what I remember him saying:

1. On the first day of classes, there was no orientation for new teachers, no brief meeting where the Principal shook his hand and said “Welcome Aboard,” nothing. He had to go to the front office and ask a secretary what classroom was his and walk there by himself.
2. Unexcused absences were chronic and undermined his ability to teach anything. At the start of each of his classes, he had a written roster of students, and he had to check off which students were there. For any class, typically 20-30% of students would be missing, without explanation (This is a very important point to remember whenever anyone tries to blame DCPS’ poor outcomes on large class sizes—on paper, each class might have 35 students, but typically, only 23 are actually showing up). Additionally, the 20-30% of students who were absent each class varied from day-to-day, meaning one student didn’t know what was taught on Monday, the one next to him was there Monday but not Tuesday, the third was there the first two days but not Wednesday, etc.
3. Student misbehavior was atrocious. For example, out of the students who showed up to class, it was common for some to walk into the classroom late, again without any explanation and often behaving disruptively. As a rule, whenever a student did that, he was obligated to sign his name on a clipboard for the teacher’s attendance records (there was no punishment for tardiness—late students merely had to write their names down). Some late students would chronically resist doing this, either ignoring him and just going to their desks or yelling curses at him. My friend described an incident where one student—who was physically bigger than he was—yelled out he was a “FAGGOT” when asked to sign the clipboard, provoking laughs from all the other students, before sitting down without signing it. After seeing he could get away with that, the student started calling my friend “FAGGOT” all the time. Other examples of misbehavior included near-constant talking among the students during lessons and fooling around with cell phones.
4. Teachers received almost no support from the school administration. Had sane rules been followed at this high school, students would have been immediately sent to the office for formal punishment for these sorts of offenses I’ve described. However, under such a policy, the office would have been overwhelmed with misbehaving students and probably some of their enraged parents, so the administration solved the problem by forbidding teachers from sending students to the office for anything other than physical violence in the classroom. My friend had no ability to formally punish the student who liked to call him “FAGGOT” other than to use stern verbal warnings.
5. Most of the students were unwilling and in some cases unable to learn. During class sessions, the students were clearly disengaged from what he was teaching. Homework completion rates were abysmal. As the end of the academic semester neared, he saw that a huge fraction of them were on track to fail, so he resorted to pitiful cajoling, pizza parties, reward schemes, and deals involving large curves to everyone’s grades if they could only, for once do a little work, and it didn’t work. Some of his students were Latino and understood little or even no English, meaning they learned (almost) nothing, even when they tried. He resorted to seating the students who knew no English next to bilingual Latinos who could translate for them. That was the best he could do. In fairness, he spoke glowingly of some of his students, who actually put in some effort and were surprisingly smart [...]

I’ll never forget how crestfallen and stressed out he was when he described these things to me. Having never taught in American public schools, I didn’t realize just how bad it was, and the detailed nature of his anecdotes really had an impact on me. I advised him to finish his year at the high school and then to transfer to ANY non-urban school in the area, even if it meant lower pay or a longer commute. We lost touch after that, but I can’t imagine he still works in DCPS.

The education system remains popular because they can always hold up glossy posters of smiling upper-class children at Rich Oaks Magnet High School and claim the system works. But basic jobs are going to be selecting primarily from

the very poor demographic and they're going to get hit with the same problem as the poorest public schools – a need for people to behave, combined with inability to credibly disincentivize misbehavior.

Basic income avoids this problem. It provides money to everyone, good employees and bad employees alike, without forcing any workplace to keep people it finds unproductive or threatening, and without having to find humiliating make-work jobs for anybody.

7. Private employees deal with bad workplaces by quitting them; nobody has a good plan for how basic jobs would replace this

And if you think this is a problem for the managers, just wait until you see what the employees have to put up with.

Some bosses are incompetent. Some are greedy. Some are downright abusive. Some don't have any obvious flaw you can put your finger on, they just turn every single day into a miserable emotional grind. Sometimes the boss is fine, but the coworkers are creeps, or bullies, or don't do their fair share. Sometimes the boss and the coworkers are both okay, but the job itself just isn't suited to your personality and what you can manage.

In private industry, people cope by leaving their job and finding a better one. It's not a perfect system. A lot of people are stuck in jobs they don't like because they're not sure they can find another, or because they don't have enough money to last them through the interim. And this is one reason why poor people who can't easily change jobs have worse working conditions than wealthier people who can. But everyone at least has the option in principle if their job becomes unbearable.

What about the people who can't get any jobs besides the guaranteed basic ones? How do they deal with abusive working conditions?

Probably somebody will set up some system to let you quit one basic job and go to a different one in the same city. But probably it will end up being much more complicated than that. How do you deal with the guy who quits every job after a week or two, looking for the perfect cushy position? How do you deal with the case where there's only one basic job available within a hundred miles? How do you deal with the case where everyone wants the same few really good jobs, and nobody wants to work at the awful abusive soup kitchen down the road?

People will set up systems to solve these problems, and the systems will be unwieldy and ineffective, just like the systems for switching public schools today, and just like all the other clever top-down socialist systems people invent to replace exit rights. Probably they'll take the edge off some of these problems, but probably nobody will be truly satisfied with the results.

Basic income solves this problem. It doesn't make anybody stay at a workplace they don't like.

8. Basic income could fix private industry; basic jobs could destroy it

In my dreams, the government finds a way to provide a basic income at somewhere above subsistence level. The next day, every single person working an awful McJob quits, because there's no reason to work there except not being able to subsist otherwise.

After that, one of two things happens. First, maybe McDonald's makes a desperate effort to invent awesome robots that can serve food without human support. Society and Ronald McDonald share a drink together – McDonald's has managed to remain a profitable company providing a valuable service, and poor people live comfortable lives without having to flip burgers eight hours a day.

Or maybe inventing robots is hard, and McDonald's has to lure some people back. They raise pay and improve working conditions, until the prospect of working for McDonald's and getting luxuries is better than the prospect of living off basic income and getting subsistence. Maybe McDonald's has to raise prices; maybe they even have to close some stores. But again, something like McDonald's continues to exist and workers are relatively well-off.

A poorly-planned basic jobs guarantee could make the problem worse. Suppose that the government decided to use its free labor to farm cows. This puts various private cow-farming companies out of business; after all, the government can pay its employees out of the welfare budget, but private companies have to pay employees out of revenue. Some of the unemployed cow-farmers go get a guaranteed basic job, putting further private companies out of work. And other unemployed cow-farmers go work at McDonald's, driving up the supply of McDonald's employees and so ensuring lower wages and worse conditions.

This isn't to deny that a well-planned basic jobs guarantee could have the same effect as basic income; if the government jobs were better than McDonald's, McDonald's might have to raise wages and improve conditions to lure people back. The direction of the effect would depend on how good the government jobs are and how much they compete with private industry. I predict the government jobs will be very bad, and compete with private industry a lot, which makes me expect the effect will be negative.

9. Basic income supports personal development; basic jobs prevent it

I have a friend who was stuck on a dead-end career path. His job paid a decent amount, he just didn't really like where it was going. So he saved up enough money to live on for a year, spent a year teaching himself coding, applied to a programming job, got it, and felt a lot more comfortable with his financial situation.

And I had a patient in a similar situation. Hated her job, really wanted to leave it, didn't have enough skills to get anything else. So she went to night school, and – she found she couldn't do it. After working 8 to 6 every day, her ability to go straight from a long day's work to a long night's studying just wasn't in the cards. And her income didn't give her the same opportunity to save up some money and take a year off. So she gave up and she still works at the job she hates. The end.

Basic income would give everyone who wants to work the same opportunity as my friend – the ability to take a year off, cultivate yourself, learn stuff, go to school, build your resume – without it being a financial disaster.

Basic jobs would leave everyone in the same position as my patient – forced to work 40+ hours a week, commute however many hours a week, good luck finding time to earn yourself a ticket out of that lifestyle while still staying sane.

There are more creative things you can do with time off work. Entrepreneurs like to talk about “runway” – how long can you keep burning through money before you run out and have to declare your new business a failure? Sometimes your runway is costs like renting an office or paying employees, but for small one-person businesses the question is usually “how long can I continue to live and feed myself working on this not-yet-profitable company?”

And poor people have runway issues of their own. One of the most common reasons poor people end up in crappy jobs is because they don't have the luxury of a long job search. If your savings will only last you a month before you can't make rent, you're going to accept the first job that will take you and feel grateful for it. If you have a guaranteed income source, you can wait until somebody presents you with a better fit.

Basic income is unlimited runway. Entrepreneurs can feel free to try out crazy ideas without the constant pressure of losing their shirt; people in between jobs can feel free to spend time looking for options they can tolerate.

Basic jobs solves none of these problems, and maintains the time pressures that prevent people from exploring interesting ideas or realizing their full potential.

10. Basic income puts everyone on the same side; basic jobs preserve the poor-vs-the-rest-of-us dichotomy

Welfare users often talk about the stigma involved in getting welfare. Either other people make them feel like a parasite, or they just worry about it themselves. Basic jobs would be little different. There will be the well-off people with jobs producing useful goods and services. And there will be the people on guaranteed basic jobs, who know their paychecks are being subsidized by Society. In the worst case scenario, people complaining about workplace abuses at their guaranteed basic job will be told how lucky they are to have work at all.

Basic income breaks through that dichotomy. Everybody, from Warren Buffett to the lowliest beggar on the street, gets the same basic income. We assume Warren Buffett pays enough taxes that the program is a net negative for him, but taxes are complicated and this is hard to notice. Rich people are well aware they contribute more to the system than they get out. But they don't think of it on the level of "I pay \$340 in taxes to support my local police station, but only get \$154.50 of police services. Meanwhile, Joe over there pays \$80 in police taxes and gets \$190 in police services. I hate him so much!"

There will be people on basic income who have no other source of money. There will be people who supplement it with odd jobs now and then. There will be people who work part-time but who plausibly still get more than they pay in taxes. There will be people who work full-time and maybe pay more than they get but aren't really sure. At no point does a clear dichotomy between "those people getting welfare" and "the rest of us who support them" ever kick in.

11. Work sucks

Amidst all of these very specific complaints, I worry we're losing site of the bigger picture, which is that work sucks. I have my dream job, the job I've been lustng after since I was ten years old, it's going exactly as well as I expected – but I still Thank God It's Friday just like everyone else.

And other people have it almost arbitrarily worse. Here are some of the cases you hear about several times a week doing psychiatry:

"I work really long days at my job. I have to deal with angry clients, bosses who don't appreciate me, and coworkers who try to dump their work on me. By the time I get home after my hour-long commute, I'm too wiped to do anything other than make a microwave dinner and watch TV for an hour or two until I pass out. Then on the weekends I take care of business like grocery shopping, cleaning, and paying my bills. Then Monday comes around and I have to do it all over again. I feel like work drains all my energy and doesn't leave me any time to be me. I used to play in a band, and we had dreams of making it big, but I had to quit because I don't feel like I have time for it any more. It's just work, go home, sleep, repeat."

"I can't stand the new open office plan. I feel like I've got to do work in the middle of a loud bar where everyone's trying to talk over each other. Sometimes I hide in the janitorial closet just so I can concentrate for a couple of hours while I finish sometimes important. I'm afraid if anyone ever catches me doing that they'll say I'm 'not a team player' and I'll get written up, but I just can't take being crammed together with all those people. Maybe if you gave me some Adderall I could focus better?"

"Sorry I haven't seen you in a few months. My workplace says it gives time off for doctor's appointments, but you still get in trouble for missing targets, and I just couldn't find any time that works. I ran out of my medication a month ago

and am having constant panic attacks, so if you could refill that right away it would be nice. And sorry, I need to go now, I'm actually calling you from the bathroom. I wanted to call you from the janitorial closet, but when I went in, there was a woman inside who mumbled something about the open office plan and accused me of distracting her."

And the people with the *worst* jobs don't have good enough time or money to see psychiatrists; I just never meet them. But I understand it gets pretty bad:

Amazon employee here. The post [[The Undercover Author Who Discovered Amazon Warehouse Workers Were Peeing In Bottles Tells Us The Culture Was Like A Prison](#)] is pretty spot on. They don't monitor bathroom breaks, but your individual rate (or production goal) doesn't account for bathroom breaks. Or let's say there is a problem like you need two of something and there's only one left, well you have to put on your "andon", wait for someone to come "fix" for you, all the while your rate is dropping. The two most common reasons people get fired are not hitting rate, and attendance. They don't really try to help you hit rate, they just fire and replace.

My first week there two people collapsed from dehydration. It's so common place to see someone collapse that nobody is even shocked anymore. You'll just hear a manager complain that he has to do some report now, while a couple of new people try to help the guy (veterans won't risk helping because it drips rate). No sitting allowed, and there's nowhere to sit anywhere except the break rooms. Before the robots (they call them kivas) pickers would regularly walk 10-15 miles a day, now it's just stand for 10-12 hours a day.

People complain about the heat all the time but we just get told 80 degrees (Fahrenheit obviously) is a safe working temp. Sometimes they will pull out a thermometer, but even when it hits 85 they just say it's fine.

There's been deaths, at least one in my building... Amazon likes to keep it all hush hush. Heard about others, you can find the stories if you search for it, but Amazon does a good job burying it.

Every now and we have an inspection, where stuff like this should be caught and changed. But they just pretty it up. If the people doing the inspection looked at numbers on inspection day vs normal operation, they would see a massive difference... but no fucks given.

The truth is the warehouses operate at a loss most the time, Amazon literally can't afford to pay the workers decent pay, and can't afford to not work them to death. The entire business model is dependent on cheap (easily replaceable) labor, which is why tier 1s are the bulk of the Amazon work force. My building has like 3-5k workers most the time and around 10-30k on the holiday (what they call peak). Almost all of that is tier 1, most states have 4-7 of these warehouses, and some like Texas and Arizona have tons more.

Next time you order something off Amazon, remember it was put in that box by a guy sweating his ass off trying to put 100-250 things in a box per hour, for 10 hours a day or he will be fired, making about a dollar more than minimum wage. Might have even been a night shift guy, who goes to work at 630pm and gets off at 5am.

I 100% understand that advocates of basic jobs insist that they'll be better than that, that they guarantee really good jobs in clean sunny offices where everybody has a smile in their face and is well-paid. I also understand they said the same thing about those DC public schools before throwing huge amounts of money at them. Forget promises; I care about incentives.

Either one of basic jobs or basic income could be potentially the costliest project the US government has ever attempted. Government projects usually end up cash-constrained, and the costliest one ever won't be the exception. The pressure to cut corners will get overwhelming. It's hard to cut corners on basic income – either citizens get their

checks or they don't. It's simple to cut corners on basic jobs. You do it the same way Amazon does – you let working conditions degrade to intolerable levels. What are your workers going to do? *Quit*? Neither Amazon nor government-guaranteed basic jobs need to worry about that – both know that their employees have no good alternatives.

Gathering a bunch of disempowered poor people in a place they're not allowed to opt out of, with budget constraints on the whole enterprise, is basically the perfect recipe for ensuring miserable conditions. I refuse to believe that they will be much better than private industry; the best we can hope for is that they end up no worse. But the conditions in private industry are miserable, even for people with better resources and coping opportunities than basic jobs recipients are likely to have.

I grudgingly forgive capitalism the misery it causes, because it's the engine that lifts countries out of poverty. It's a precondition for a free and prosperous society; attempts to overthrow it have so consistently led to poverty, tyranny, or genocide that we no longer believe its proponents' earnest oaths that *this time* they've got it right. For right now, there's no good alternative.

But if we have a basic jobs guarantee, it will cause all the same misery, and I won't forgive it. The flimsy justifications we can think up won't be up to the task of justifying the vast suffering it will cause. We can't excuse it as necessary to produce the goods and services we rely on. We can't excuse it as a necessary condition for political freedom. If a worker asks "why?", our only answer will be "because Cory Booker thought a basic jobs guarantee would play better among the electorate than basic income, now get back to packing boxes and collapsing from dehydration". There *will* be an alternative: a basic income guarantee. We will have rejected it.

I feel like as a quasi-libertarian, I sometimes downplay how awful private industry, capitalism, and the modern workplace are. If so, I apologize. The only possible excuse for defending such a flood of misery is what inevitably happens when people meddle with it. But the price of such morally tenuous greater-good style reasoning is that you need to stay hyper-aware of times when you *don't* need to defend the system, when there *is* a chance to do better without destroying everything. I think basic income is such a chance. And I think basic jobs are a tiny modification to the idea, which destroys its potential and perpetuates all the worst parts of the existing system.

*
*

It would be unfair to make this argument without responding to jobs' proponents' points, so I want to explain why I don't think they provide a strong enough argument against. These will be from [the Sarris piece](#). I don't want to knock it too much, because it's a really fair and well-written piece that presents the case for jobs about as well as it can be presented, and any snark I might give it below is totally undeserved and due to personal viciousness. But it argues:

i. Studies of UBI haven't been very good, so we can't know if it works.

Studying a UBI pilot with an end date is not studying UBI at all: It is instead studying a misnamed temporary cash payment. By the nature of pilots, the cohort's behavior cannot reliably change to depend on UBI's long term existence. No study yet has guaranteed a cohort money forever, and even if it did it would be difficult for a pilot to study the long term effects, some of which may be generations out. What pilot can tell us what it's like for kids to grow up with parents who have never worked? [...]

Basic Job programs are more amenable to piloting and a gradual roll-out, since new clusters of jobs appear (and end) all the time. Piloting Basic Jobs can be tried in different communities with varying magnitudes. The legislation to justify such a pilot may already be in place[1], and even a pilot may have lasting benefits. What we learn from the pilot will be more applicable than studying temporary cash transfers in a community and expecting that knowledge to translate into society-wide UBI. If a pilot is successful, one can imagine a kind of National Civil Service, organized like existing federal programs such as the National Park

■ Service, which can hire professionals to train and supervise projects.

I have some minor caveats – Alaska has had a (very small) universal basic income for some time, which seems to have worked relatively well. And basic job studies will also have trouble scaling; smaller trials might preferentially select the most functional unemployed people, would have less impact on private industry, and can always just dismiss people back to the general pool of the unemployed. But overall I agree with the point that basic income is a bigger change and we should be more suspicious of bigger changes.

But at some point you're arguing against testing something because it's untested. If we can't 100% believe the results of small studies – and I agree that we can't – our two options are to give up and never do anything that hasn't already been done, or to occasionally take the leap towards larger studies. I think basic income is promising enough that we need to pursue the second. Sarris has already suggested he won't trust anything that's less than permanent and widespread, so let's do an experiment that's permanent and widespread.

ii. UBI gives everyone the same amount, but some people need more (for example, diabetics need more money to pay for insulin). Existing social programs like medical aid take this into account; UBI wouldn't.

This seems like exactly the problem that insurance exists to solve. Bringing insurance into the picture, "everybody has to get this" switches from a negative to a positive.

I won't speculate on how this will look, except to note that it would work well with some kind of mandate where the cost of a Medicare-like state insurance gets auto-deducted from your UBI. Since I'm quasi-libertarian, I would support people's right to opt out of this, after signing and notarizing a bunch of forms with "I UNDERSTAND I AM AN IDIOT AND MIGHT DIE" on them in big red letters, but I understand other people might prefer to avoid the chance of moral hazard. It still seems like this problem is solvable.

iii. Somehow even if everyone has more money they won't be better off

■ One of the biggest assumptions people make with UBI is that the problems of today and the near future are primarily ones of money. I don't think the data supports this. [link to various charts showing that people generally have food and access to health care]

On some level, if you're tempted to believe this you should find a poor person and ask them how they feel about being poor. I predict they will say it is bad. They will not agree that our society has basically solved all of its money-related problems. They will say there is a very real sense in which their money-related problems remain unsolved. I guarantee you they will have very strong feelings about this.

But that's overly pat. A steelman of Sarris' point might go something like this: it definitely seems true that there is some complicated way in which a family of eight living in a tiny farmhouse in the Kansas prairie in 1870 was happy and felt financially secure even though they probably only earned a few hundred dollars a year by today's measures. So isn't it weird that people earning twenty thousand dollars a year still think of material goods as their barrier to happiness?

I think explaining that effectively would require a book-length treatment. But I think the book would end with "even though it's weird and complicated, poor people today who make \$10,000 or \$20,000 are often unhappy, in a way that richer people today aren't, and this involves money in a real sense."

I am not the person to write this book (though see the post on [cost disease](#)); I can only relay what poor people tell me. Sometimes it's "my rent-controlled apartment is underneath noisy frat boys who keep me awake every night with their parties, but I can never leave because it's the only apartment I can afford in this town." Sometimes it's "I hate my boss but I can't leave because if I go a month without getting a paycheck I won't have enough money for rent." Sometimes it's "I couldn't afford good birth control, got pregnant, and now I can't afford to support the child, what do I do?" Sometimes it's "Obamacare mandates me to buy health insurance, but I can't afford it, I guess I am going to have to pay a fee I can't afford on tax day instead." Sometimes it's any of a thousand versions of "my car broke down and I can't afford to get it fixed but I need to get to work somehow". Sometimes it's "I am sick but if I miss a day of work my company will fire me, because when you're poor enough legally-enshrined workplace protections somehow fail to exist in real life". And sometimes it's "I work eighty hours a week driving for Uber because it's the only way to make ends meet, I hate everything." A lot of times it involves the same crappy job-centered lifestyle I worry a basic jobs guarantee would perpetuate forever.

Trying to steelman the "it's not money" point further takes us to [Sarris' other essay on UBI](#), where he writes:

Rent is currently eating the world. Rental income just hit an all-time high. If everyone is given a very predictable amount of money, it may be seen as a system that can be gamed by landlords and maybe other essentials producers. Implementing UBI without reforming land use and zoning regulations may end up as nothing more than a slow transfer to landlords. What are the odds of that happening? Well, it seems like it already did happen with healthcare and college tuition (loans) in the US, and if those are our guide, the "money" part and the "meaningful reforms" part should be done in a very particular order.

Since housing does work well in some places (Japan and Montreal come to mind) I think this is a problem that can be fixed. But without the fix first, UBI may be punting real political problems while giving the appearance of solving them (until years later), and making the price inflation obvious for landlords, just like it was for healthcare companies and colleges getting guaranteed loans.

Payments as a solution to a broken system is not the same as fixing the system. If UBI punts this real problem, we'll be creating a financial time bomb.

This is basically how I think about any request for giving more money to education or health care, so I guess I have to take it seriously. Maybe the situations aren't exactly the same – education and health care seem to eat up money by hiring administrators, which doesn't have an obvious analogy to ordinary individuals. But the Kansas farmhouse example suggests that something like this must go on even at the personal level.

It looks like probably what's being described is that – absent some magical ability to create new houses out of thin air (a task known to be beyond the limits of modern [technology](#)) – housing is a positional good and so raising the position of everyone equally will just give extra cash to landlords. The best that can be said here is that insofar as these goods aren't perfectly inelastic, basic income will help a little. And insofar as other goods used by poor people (cars? furniture? generic medications?) are decently elastic, basic income will help a lot. I do agree the problem exists.

But I think this is one case where basic income is clearly better than basic jobs. All basic jobs can do is give you money, which can get eaten by rent-seekers. Basic income gives you freedom. Somebody works 50 hours a week at two McJobs to afford an apartment, gets basic income, and then they work 20 hours a week at one McJob and afford their apartment. The price of an apartment doesn't change, but their life has improved.

And by lowering the demand for jobs, basic income provides the seed of a solution to the housing problem. The reason rent costs so much in the Bay Area is because everyone wants to live in the Bay Area because it has so many great jobs. You can buy a house in the country (or in an unpopular city) for cheap; people don't because the jobs aren't as good, or the good jobs take longer to find. Freed from the need to live right in city center (or right next to the subway stop

leading to city center), people can spread out again. If rent is \$2000 in San Francisco and \$500 in Walnut Creek, they can live in Walnut Creek (and still go to San Francisco whenever they want – cities are very accessible from suburbs, for every purpose *except* commuting during rush hour five days a week).

Go to the suburbs and people are building new housing tracts all the time. Supply is elastic and everyone's backyards are so far away from one another that NIMBYs mostly stay quiet. It's only when our job-centered culture forces everybody into historic San Francisco city center that we start having problems.

There's still going to have to be a hard battle against cost disease. But much of the cost disease comes from overregulation and creeping socialism, and much of overregulation and creeping socialism come from well-intentioned concerns about the poor. Witness how California's recent housing bill was opposed by socialists making vague warnings about "greedy developers". If we can solve the non cost-disease-related parts of poverty first, maybe the socialists will lose some power and we can start fighting the cost disease problem in earnest.

iv. Without work, people will gradually lose meaning from their lives and become miserable

After claiming that money isn't really a problem for most people, Sarris continues:

The biggest societal ill today is not that people don't have enough money to survive, it is that to survive and thrive people need things beyond food and rent: Social responsibility, sense of purpose, community, meaningful ways to spend their time, nutrition education, and so on. If we fixate merely on the money aspect, we may be misdiagnosing what is making our 21st century so miserable for so many people.

From some psychologists' points of view, one of the worst things you can do to someone who is suffering from addiction or loss of hope is to give them no-strings-attached money, when what they really need is regularity and the responsibility that comes from having a purpose, even if it's simply a job or a station. Basic Jobs have a chance of making the opioid crisis better, UBI risks making it worse... the at-risk population in the US need functions and responsibility more than just a check.

Social responsibility. Sense of purpose. Community. Meaningful ways to spend your time. This is some big talk for promoting jobs that in real life are probably going to involve a lot of "Do you want fries with that?" Getting a sense of purpose from your job is a crapshoot at best. Getting a sense of purpose outside your job is a natural part of the human condition. The old joke goes that nobody says on their deathbed "I wish I'd spent more time at the office", but the basic jobs argument seems to worry about exactly that.

And let's make the hidden step in this argument explicit. Everyone on basic income will have the opportunity to work if they want. In fact, they'll have more opportunity, since people who hate working will have dropped out of the workforce and demand for labor will rise. So the basic jobs argument isn't just that people need and enjoy work. The argument is that people need and enjoy work, but also, they are too unaware to realize this, and will never get the work they secretly crave unless we force them into it.

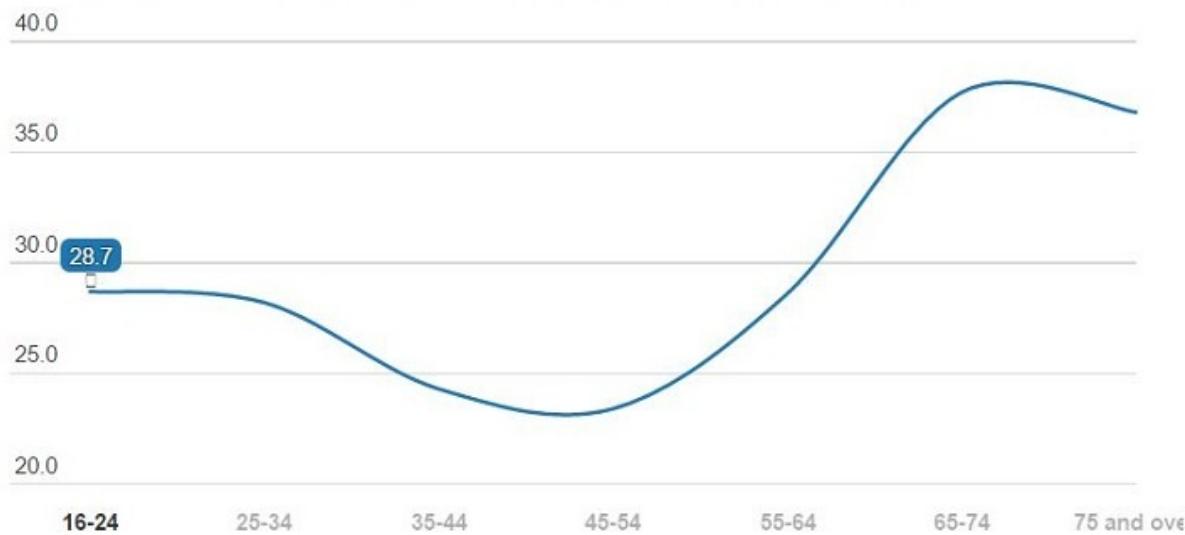
That doesn't seem right. I don't know enough hopeless opiate addicts to contradict an apparent psychological consensus on them, but it seems to me a lot of people do perfectly well finding meaning on their own time.

What about the retired?

The graph of happiness vs. age looks like this:

The happiness U-shape

Overall life satisfaction in the UK: self-reports from four surveys between April 2014 and March 2015.



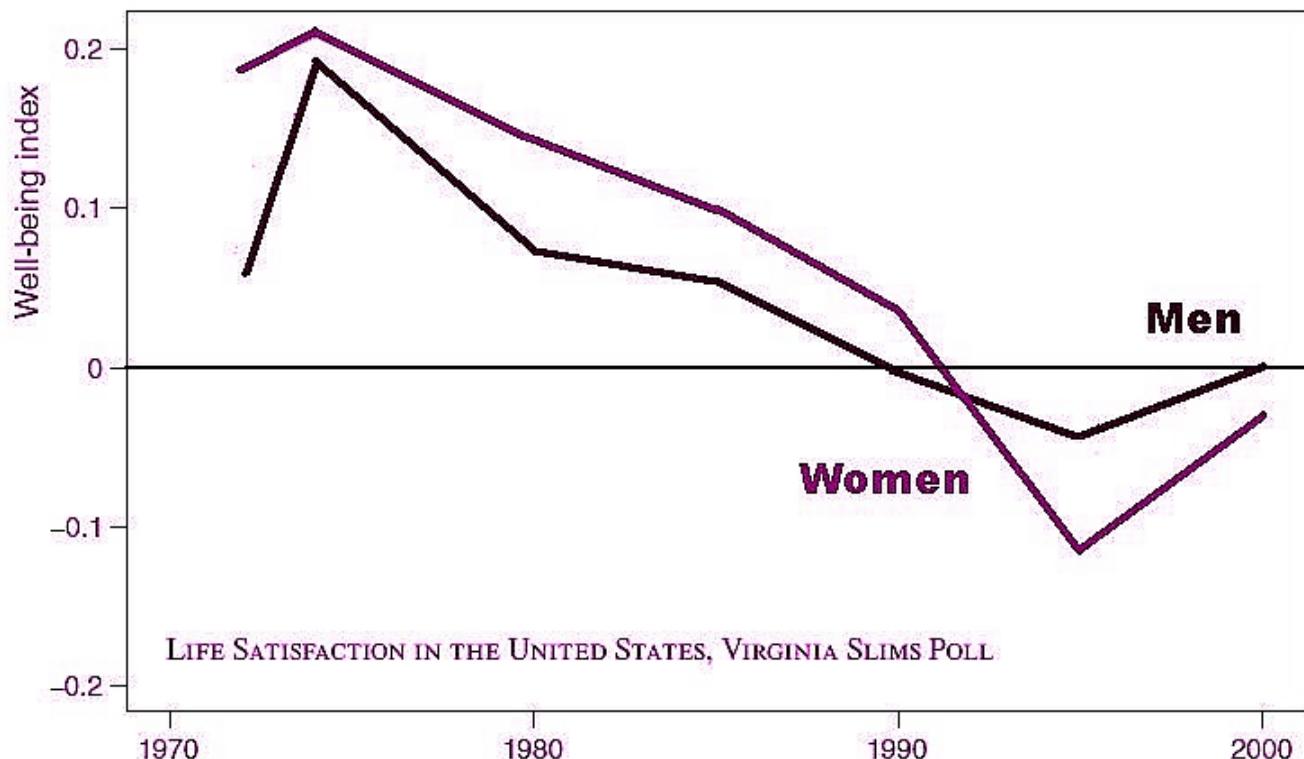
Source: UK Office of National Statistics [Get the data](#)

This is not the shape we would expect if stopping work suddenly made you miserable and deprived you of purpose. Retired people seem to avoid work just fine and have lots of fun golfing, watching golf tournaments, going on golf vacations, arguing about golf, and whatever else it is retired people do.

Sarris says that "If you think UBI would not make the opioid crisis worse, the onus is on UBI proponents to show how writing 'UBI' on the top of the check instead of 'disability' would do that." I would counter-argue that the onus is on opponents to explain why writing 'UBI' on the check works so much worse than writing 'Social Security'.

What about homemakers?

Yes, homemaker is a full-time job. But it's the full-time job a lot of people would do if they didn't have to do their regular full-time job, which makes it fair game when we're talking about basic income. Here's a graph of male vs. female happiness over time:

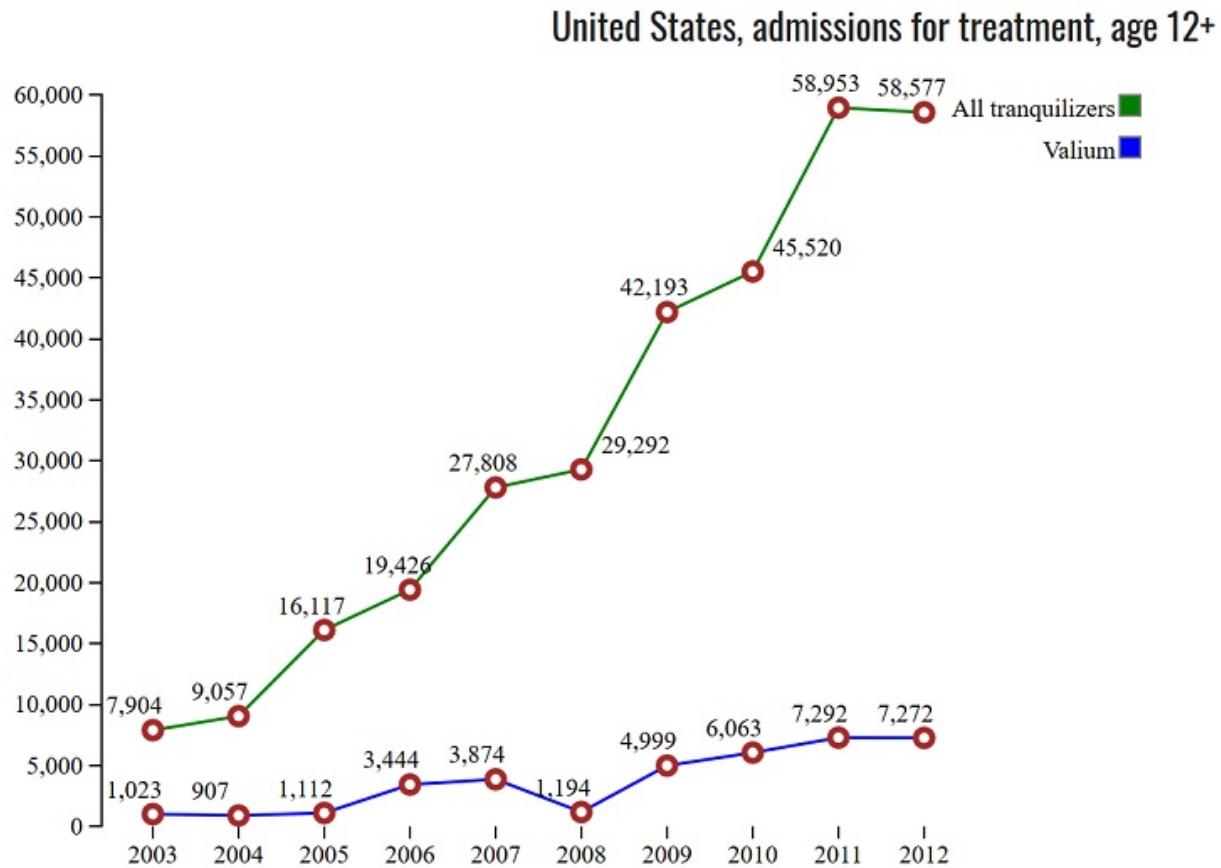


If we assume most women in 1970 were homemakers, and most women in 2000 are working, their shift from homemaking to working doesn't correspond to any improvement in happiness, either absolutely or relative to men.

There is some debate over whether modern-day homemakers are happier than modern-day workers or vice versa, with the most careful takes usually coming down to "people who prefer to stay home are happier staying home, people who prefer to work are happier working". But there is no sign of the collapse in meaning and happiness we would expect in homemakers if not having an outside-the-house job reduces you to purposeless nihilism.

When I bring this up to people, they always have the same objection: "Didn't women back then use lots of tranquilizers because of how stressed and upset they were? Didn't they even call Valium 'Mother's Little Helper?'" Yes. But take it from a psychiatrist who prescribes them: people still use lots of tranquilizers. Nobody cares anymore, because it's no longer surprising or ironic.

Treatment facility admissions



Sure glad that tranquilizer overuse problem got nipped in the bud in the 1970s when we cancelled stay-at-home parenting.

What about aristocrats?

History presents us with many examples of entire classes who managed to live off other people's work and avoid working themselves. These people seem to have not only have been pretty happy with the deal, but often used their free time to contribute in less purely economic ways. Lord Byron and Warner von Braun were hereditary barons, Bertrand Russell a hereditary earl, de Broglie a hereditary Duke, Condorcet and de Sade hereditary Marquises. Von Neumann's family was some kind of nouveau riche Austro-Hungarian nobility; Wittgenstein's family was something similar. Winston Churchill was grandson of a Duke and son of a Lord. None of them ever had to worry about money: society gave them a giant basic income check from their ancestral estates.

Yet Churchill found meaning by saving the UK. Von Braun found meaning by shooting missiles at the UK. Condorcet found meaning by becoming one of the foremost defenders of human rights. De Sade found meaning by becoming one of the foremost violators of human rights. De Broglie and von Neumann found meaning by contributing to fundamental physics. Russell and Wittgenstein found meaning by literally figuring out what meaning was. Overall they seem like a pretty flourishing bunch.

What about college students?

Technically they have to go to classes, but a lot of them get away with ten hours or less of class per week, and even

more of them just never attend. Some, like Mark Zuckerberg and Bill Gates, use the extra time to found startups. Others, like everyone else, use the extra time to party and take lots of drugs. Either way, they seem pretty happy.

What about the self-employed?

Being self-employed costs you a lot of the supposed psychological benefits of work. You might not be leaving the house. You might not be interacting with other people. But [studies find that](#) the self-employed are happier than the other-employed, even though they work longer hours and have less job security.

What about hunter-gatherers?

Hunting-gathering in a fertile area is a pretty good gig, and usually lets people support themselves with only [a few hours' work per day](#). Most evidence suggests they're [pretty happy](#) despite their lack of material goods.

What about schoolchildren?

Every year, I would complain that I hated school. Every year, my mother would repeat some platitude like "Oh, when summer comes around you're going to be so bored that you'll be begging to go back". And every year, summer vacation would be amazing, and I would love it, and I would hate going back to school with every fiber of my being. I understand this is pretty much a consensus position among schoolchildren. This has left me forever skeptical of arguments of the form "Oh, if you had freedom you would hate it".

What about me?

When I graduated medical school, I applied to residency and was rejected. That left me with a year open before I could try again. Thanks to some odd jobs, a little savings, and charity from friends and family, I was able to subsist. I spent the year meeting new people, hiking around California, falling in love, studying philosophy, and starting this blog. At the end of the year I applied to residency again and was accepted. I'm glad I got the job I wanted, but I also remember that year fondly as maybe the best I've ever had, and the one that set the stage for a lot of the good things in my life that happened since. I think this is pretty common for well-off people. We call it a "gap year" or a "sabbatical" or "going off to find yourself" or any of a bunch of other terms that disguise how it's about doing exactly what people say you can't do – being happy without a 9 to 5 job.

When I bring these points up, basic jobs advocates usually find reasons to dismiss all of them. Schoolchildren and college students are at a special part of their life that doesn't generalize. Homemakers like being with their kids. Aristocrats get the world as their oyster. Retirees are mysteriously and permanently mesmerized by golf, which becomes an ur-need subsuming all other human desires. Hunter-gatherers are evolutionarily adapted to their lifestyles. I am just weird. They dismiss all of these as irrelevant and go back to their core example: in the US, right now, unemployed and disabled people are terribly unhappy.

I accept the very many studies that show this, but I do wonder if this has more to do with contingent features of unemployment than with work being necessary to human flourishing. For example, unemployed people are chronically low on money. Unemployed people face stigma and constant social pressure to get employment. Unemployed people live in a society built around and emphasizing jobs. Unemployed people may have pre-existing problems in their lives that led to their unemployment. Unemployed people sometimes suffer from disabilities or chronic pain. Unemployed people have no friends to hang out with during business hours because everyone else is working.

If you compared gay vs. straight happiness in 1980, you probably would have found gay people were much less happy. Now some studies suggest that in liberal and accepting areas, they are [as happy or happier](#). The relative happiness of different groups isn't necessarily a human universal; it can also depend on how society treats them.

Given all this, I lean in favor of thinking most people would tolerate financially secure leisure time just fine. I might be wrong. *But I am still more comfortable letting people decide for themselves.* People who try leisure and like it – or who prefer homemaking, or taking care of elderly parents, or anything else – will stay out of the workforce. People who try leisure and don't like it will apply for the new, better class of jobs that will exist once increased demand for labor has forced employers to up their standards. Or they'll go volunteer at their church. Or they'll start a nonprofit. Or they'll do something ridiculous like try to be the first person to unicycle around the world.

Or maybe the meaninglessness of modern life will start to recede. Why don't we have strong communities anymore? One reason I keep hearing from my patients is that they had lots of friends and family back home in Illinois or Virginia or wherever – but all the good jobs are in the Bay Area so now they live here and don't know anybody. My own friends have managed to set up a halfway-decent semi-intentional community in California, but only because by a happy coincidence they all work in computers and all the good computer jobs are in the Bay. Freeing people from needing to orient their entire life around where they can get a job might lead to a lot more intentional communities like mine. Or it might lead to other things we can't think of right now. A bunch of people with a lot of leisure time to throw at problems, and a bunch of people with money and a problem of meaninglessness, seems like a pretty good combination if you're looking for meaning-as-a-service.

The best studies on homemakers find that women who want to be homemakers are happier as homemakers and sadder if forced to work, and women who want to work are happier as workers and sadder if forced to stay at home. I would not be surprised if there are some people who are happiest working, and others who are happiest pursuing leisure activities. A basic income would make it easier for both groups to get what they want.

v. If something went wrong, basic jobs programs could be more gracefully wound down.

What if it doesn't work? What if we run out of jobs? Suppose a Basic Job program fails 20–30 years into the future. Maybe there's too much corruption or not enough oversight, or the political will is no longer there, or the money itself is no longer there. Contingency planning is good: No matter how much you trust the pilot, you still want an airplane with emergency exits.

If this happens, the side effects seem less severe (or even mildly positive) when contrasted with a UBI failure. So what if we accidentally fund farms, and bakeries, and furniture production, and house construction, and all sorts of small scale crafts across the country? Even in pessimistic scenarios we can expect some of the businesses and functions built to continue serving their communities after an official program is gone, in the same way that the Hoover dam is still there. A Basic Job program can plan for contingencies and the divvying up of what's been created, democratically, by community. Sheep farmers that are no longer supported by the government have at least got their flocks. If things ever go south, Basic Jobs better position us to try something else.

"So what if we accidentally fund farms?" asked Stalin, creating the [kolkhozes](#). Maybe I am being mean here, but "let's guarantee full employment by sticking poor laborers on a government farm somewhere and teaching them to till the earth" is a plan that [ought to set off as many historical alarm bells](#) as "let's do something about all the Jews around here" or "let's murder the Mongol trade delegation".

True, nobody is proposing the other prong of socialist agricultural policy, which is crushing the private farms. But it's important to remember that what's being proposed is basically socializing large parts of the economy in ways that

history tells us lead not only to agricultural catastrophe when being set up, but to economic ruin when being wound down:



In the 1990s, the GDP of Russia declined by 50%. Fifty percent! I don't know if that's ever happened before in history outside of a civil war or foreign invasion. The Iraqi economy survived the Iraq War and subsequent sectarian conflict better than the Russian economy survived winding down its basic jobs program.

Maybe I'm being unfair. Socializing part of the economy is probably safer than socializing all of it. And not crushing the private farms really does provide a safety valve that previous collectivization efforts lacked (though if the government farms are more subsidized than they are inefficient, you'll crush the private farms whether you want to or not).

But I'm still not sure if unsocializing the economy is as easy as winding down a basic income. If you want to wind down a basic income, you decrease it by 5% per year, and each year more people go to work in the private sector or start training to do so. If you want to wind down a nationwide system of collective farms, you – well, empirically you flail about for a while, collapse into a set of breakaway republics, and end up getting ruled by Vladimir Putin.

vi. Basic jobs could be used to create useful infrastructure

Have the imagination to consider all of the work that is not being done, and FDR-style public works programs can be found almost everywhere. Building bicycle lane networks. Creating and maintaining public parks, flowerbeds, sidewalks. Demolition and recycling and re-urbanization (or re-forestation) of derelict factory grounds. There are so many things that would make parts of the US better places to live. As long as swaths of America are in disrepair and also where the jobs aren't, Basic Jobs has a mission to fulfill.

Some of my concern here comes from my concern (mentioned above) that basic-job-havers would not be very good employees, and that you would probably save money by handing needy people a check and separately hiring some

super-efficient megacorporation to make your flowerbeds.

But another part comes from asking myself – which would I rather have? More flowerbeds and sidewalks? Or forty extra hours a week to spend seeing friends and family, or pursuing hobbies that I love? Framed this way, the answer is super-obvious – and remember, I love my job.

vii. Capitalism seems to have historically worked pretty well, and basic jobs guarantees preserve the best features of capitalism

We want to try and keep [the] positive effect of capitalist economic transactions. UBI creates paychecks, Basic Jobs programs do too, but Basic Jobs also create transactions, incentives, and products, fulfilling secondary needs for society.

Basic Jobs can be thought of as a program that is paying people to make other people's lives better in addition to their own. We are paying people to produce local food and crafts, in a subsidized fashion that gives communities an alternative to the WalMart-esque globalized marketplaces. If the government subsidizes the workers so that their goods can be competitive, it will foster local economies while putting money in the pockets of local workers who themselves have more power. Hopefully, the second-order effects of such commerce are large enough to notice. Maybe the benefits will stay. One could argue that the strong Swiss and other European agricultural subsidies are already a soft form of Basic Jobs.

"Capitalism" is a Rorschach test that means many things to many people. Some people think it means oppression, discrimination, and exploitation. Other people think it means any level of freedom better than you get in Maoist China. Still other people identify it with corporations, or banks, or barter, or any of a thousand other things. But to me, if capitalism means anything at all, it means...

Well, remember argument iv above? About how maybe poor people's lives will be meaningless without work, and maybe they're not sufficiently self-aware to realize that on their own, so the government should make them work for their own good, in whatever industry most needs their help?

To me, capitalism means shouting "FUCK YOU" at that argument, at the intuitions behind that argument, and at the whole social structure that makes those intuitions possible, then sterilizing the entire terrain with high-quality low-cost American-made salt so that no other argument like it can ever grow again. There are other parts of capitalism, like the stuff about stock exchanges, but they all flow from that basic urge.

Capitalism certainly doesn't mean you should never get money without working. Heck, some leftists would *define* a capitalist as a person who gets money without working. The part where you get money without working is the fun part of capitalism. The part where most people *don't* get that is the part that could do with some fixing. That's why a lot of history's greatest capitalists (in both senses of the word) – from Friedrich Hayek and Milton Friedman to Mark Zuckerberg and Elon Musk – have supported basic income.

My intuitions are basically Georgist (*note to self: read Henry George before saying this too many more times*). Capitalists deserve to keep the value they create, but they also owe rent on common resources which they enclose and monopolize (eg land, raw materials). That rent gets paid to the State (as representative of the people who are denied use of the commons) in the form of taxes. The State then redistributes it to all the people who would otherwise be able to enjoy the monopolized resources – eg everybody. I think this process where businesses pay off the government for their raw materials is pretty similar to the process where they pay off the investors for their seed money, and that the whole thing fits within capitalism pretty nicely.

I don't think the government taking a big role in the economy for Your Own Good can ever really fit within capitalism, at least not the parts of it that I consider valuable. I would consider a basic jobs guarantee, if it lasted, to be a victory for socialism over the parts of capitalism I hold dear – the final triumph of the old Soviet joke about how "We pretend to work, they pretend to pay us". If you want an image of the future, imagine a glassy-eyed DMV employee staring at a clock, counting down the hours until she can go home – forever.

And that's what we're debating here – an image of the future. These basic guarantees always get brought up in the context of technological unemployment. I've [looked into this before](#), and although I don't think jobs are being destroyed per se, I think it's definitely possible they're getting worse for complicated reasons. So as more and more people start getting worse and worse jobs, we can choose one of two paths.

First, we can force more and more people into make-work low-paying government jobs. Extrapolate to the very far future, and 99% of the population will spend their time sending their kids off to daycare before a long day of digging ditches that a machine could have dug better, while 1% of people have amazing robot empires.

Second, we can try to break the link between toiling for someone else and being able to live. We can set some tax rate and promise that all revenue above some amount necessary to fund state functions will be redistributed as basic income. It'll be pretty puny at first. But as GDP grows, more and more people will opt out of work. As the payments increase, we can gradually transfer various forms of welfare into insurance, and use the gains to grow the payments further. There will be plenty of well-paying jobs for whoever wants to keep working, and lives of leisure and enjoyment for the people who don't. Robots will pick up the slack and keep the big corporations generating the value that gets siphoned off. Extrapolate to the very far future, and 99% of people live in constantly-improving comfort and freedom, while 1% of people have that plus amazing robot empires.

Both of these are kind of tame shock-level-zero visions. But they set the stage for whatever comes next. If we have genetically enhanced superchildren, or [Hansonian em overlords](#), they're going to inherit the same social structures that were on the scene when they got here. Whatever institutions we create to contain today's disadvantaged will one day be used to contain us, when we're disadvantaged in a much more fundamental way. I want those structures to be as autonomy-promoting as possible, for my own protection.

I grudgingly admit basic jobs would be an improvement over the status quo. But I'm really scared that it becomes so entrenched that we can never move on to anything better. Can anyone honestly look at [the DC education system](#) and say "Yeah, I'm glad we designed things that way"? Doesn't matter; we're never going to get rid of it; at this point complaining about it too much would send all the wrong tribal signals. Nothing short of a civil war is going to change it in any way beyond giving it more funding. I dread waking up in fifty years and finding the same is true of basic jobs.

This is what I mean by hijacking utopia. Basic income is a real shot at utopia. Basic jobs takes that energy and idealism, and redirects it to perpetuate some of the worst parts of the current system. It's better than nothing. But not by much.

EDIT: [Sarris' response](#), where he argues that I am comparing the most utopian formulation of basic income to a very practical 'let's get a few unemployed people back to work' version of basic jobs.

Bulls**t Jobs (Part 1 of ∞)

Posted on August 29, 2018 by Scott Alexander



A surprisingly common part of my life: a patient asks me for a doctor's note for back pain or something. Usually it's a situation like their work chair hurts their back, and their work won't let them bring in their own chair unless they have a doctor's note saying they have back pain, and they have no doctor except me, and their insurance wants them to embark on a three month odyssey of phone calls and waiting lists for them to get one.

In favor of writing the note: It would take me all of five seconds. I completely believe my patients when they say their insurance is demanding the three month odyssey. Or sometimes they don't have insurance and it would be a major financial burden for them to consult another doctor. Also, I've seen these other doctors and they have [no objective test for back pain](#). 90% of the time they just have the patient stand in front of them, make whatever movement it is that hurts their back, ask the patient if it hurt their back, and when the patient says yes, the doctor says "That's back pain all right, take some aspirin or ibuprofen or whatever".

Against writing the note: I am a psychiatrist. I usually treat patients via telemedicine, which means that in many cases I have literally never seen their back. All I remember about back pain from medical school is that some people call it "lumbago", a word that stuck in my head because it sounds like a cryptid or small African nation. I know even less about the ergonomics of chairs, or when people do vs. don't require better ones. Any note I write about back pain and chair recommendations is going to be a total sham, bordering on medical fraud. I could demand my patient take time off work to come in for an examination, sometimes from several hours away, just so I can do the thing where they bend their back in front of me and tell me it hurts. But that's kind of just passing the shamminess a little bit down the line in a way that seriously inconveniences them.

In other words: the request puts me in a position where I either have to lie, or have to refuse to give people help that they really need and that it would be trivial for me to provide. It's one of my least favorite things, and I would appreciate any ethical advice the philosophers here have to give.

But my latest strategy is radical honesty. I write a note saying:

To whom it may concern:

I am a psychiatrist treating Mr. Smith. He tells me that he has chronic back pain ("lumbago"), and asks to be allowed to bring in his own chair to work.

Yours,

Dr. Alexander

It's too soon to have a good sample size. But it seems to usually work. I think it works because there is nobody at Mr. Smith's workplace – maybe nobody in the entire world – who's really invested in preventing Mr. Smith from bringing a chair into work. Someone wrote up a procedure for employees using special chairs, so that they're not the sort of cowboys who make decisions without procedures. Someone else feels like they have to enforce it, so that they're not the sort of rebel who flouts procedures. But nobody *cares*.

I think a lot about David Graeber's work on [bulls**t jobs](#). In an [efficient market](#), why would profit-focused companies employ a bunch of people who by their own admission aren't doing anything valuable? I've been wondering about this for a long time, and I try to notice when something I'm doing is bulls**t. I guess this fits the bill. It seems to be an issue of people spending time and money to create and satisfy procedures that degenerate into rituals, so that they can look all procedural and responsible in front of – courts? regulators? bosses? investors? I'm not sure. But I do wonder how much of the economy is made of things like this.

Considerations On Cost Disease

Posted on February 9, 2017 by Scott Alexander



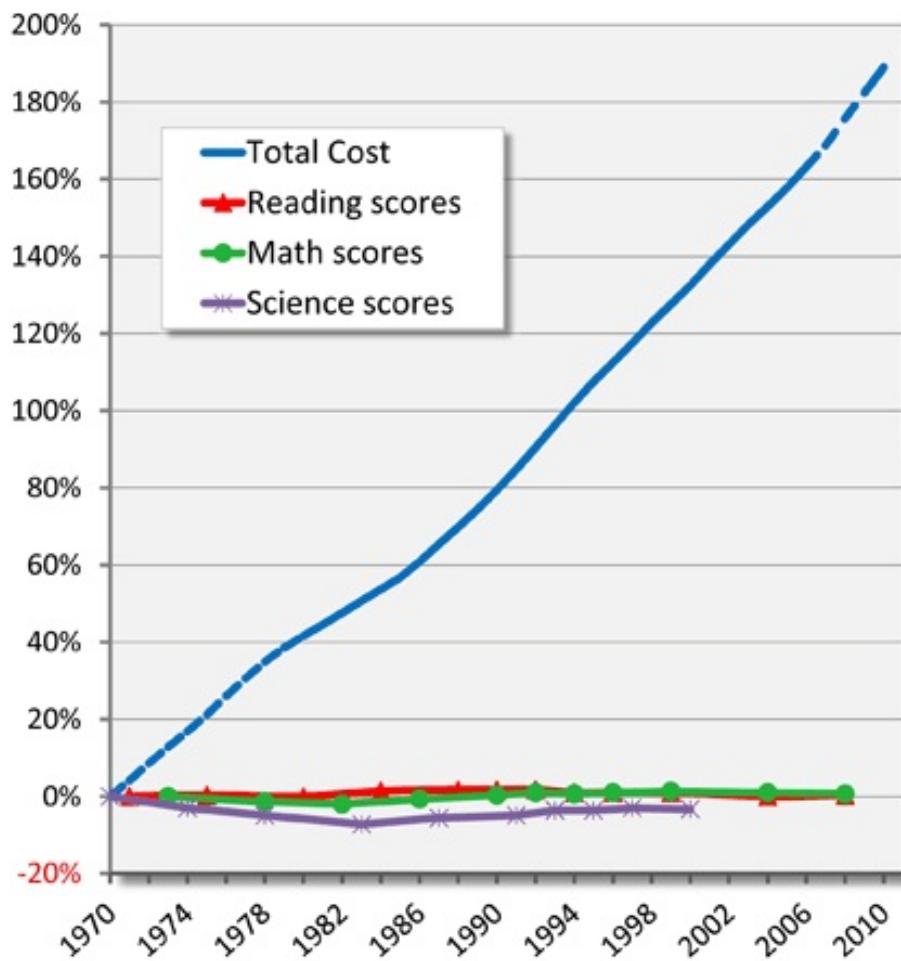
I

Tyler Cowen [writes about cost disease](#). I'd previously heard the term used to refer only to a specific theory of why costs are increasing, involving labor becoming more efficient in some areas than others. Cowen seems to use it indiscriminately to refer to increasing costs in general – which I guess is fine, goodness knows we need a word for that.

Cowen assumes his readers already understand that cost disease exists. I don't know if this is true. My impression is that most people still don't know about cost disease, or don't realize the extent of it. So I thought I would make the case for the cost disease in the sectors Tyler mentions – health care and education – plus a couple more.

First let's look at primary education:

Trends in inflation-adjusted total cost of a K-through-12 public education and achievement of 17-year-olds (U.S.)



Cato Institute

"Total cost" is the whole amount spent on the K-12 education of a student graduating in the given year. We graph the percent change in that amount--and in test scores--over time.

Data sources

(test scores): NAEP, *Long Term Trends* reports, U.S. Department of Education
(cost): *Digest of Education Statistics 2011*, Table 191, U.S. DOE, CPI adjusted to constant 2012 \$. Missing values linearly interpolated or extrapolated.

Prepared Sept. 2012 by:

Andrew J. Coulson,
Director, Cato Institute Center
for Educational Freedom

There was some argument about the style of this graph, but [as per PolitiFact](#) the basic claim is true. Per student spending has increased about 2.5x in the past forty years even after adjusting for inflation.

At the same time, test scores have stayed relatively stagnant. You can see [the full numbers here](#), but in short, high school students' reading scores went from 285 in 1971 to 287 today – a difference of 0.7%.

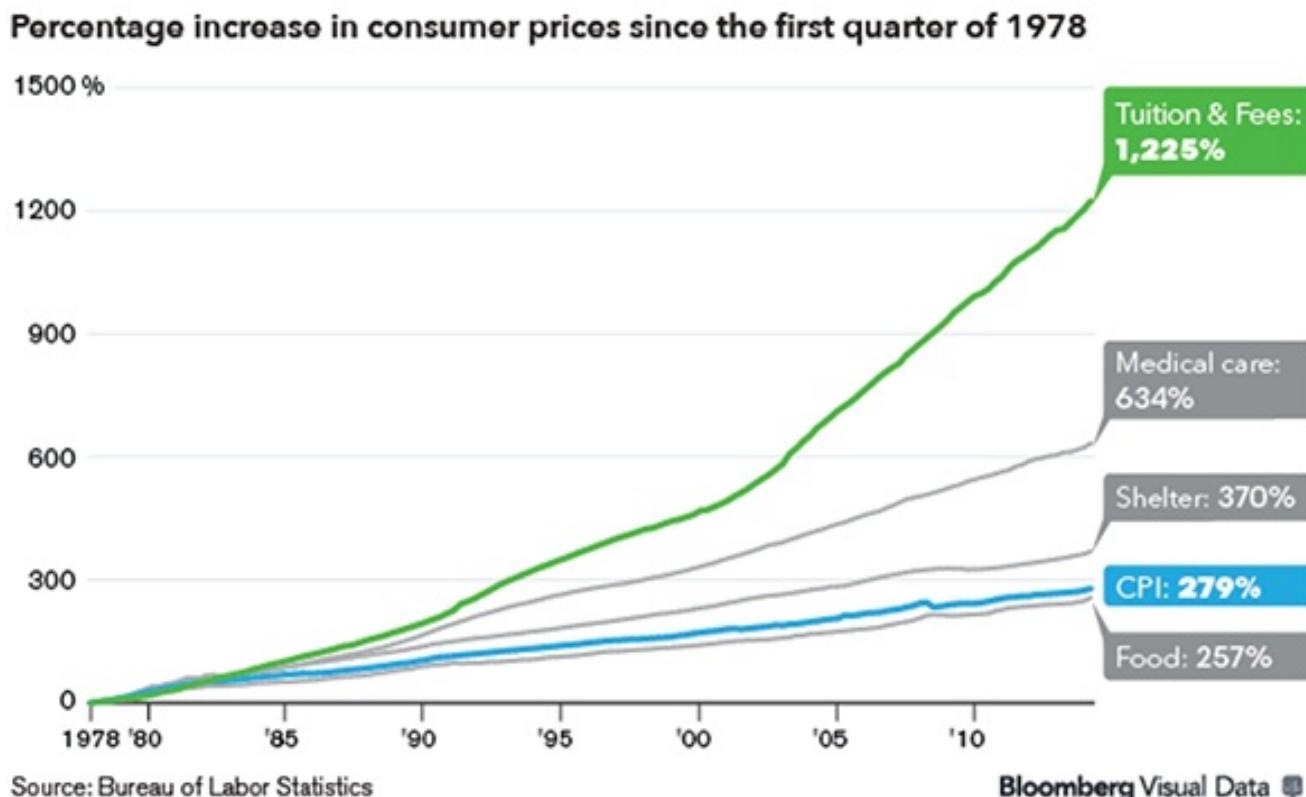
There is some heterogeneity across races – white students' test scores increased 1.4% and minority students' scores by about 20%. But it is hard to credit school spending for the minority students' improvement, which occurred almost entirely during the period from 1975-1985. School spending has been on exactly the same trajectory before and after that time, and in white and minority areas, suggesting that there was something specific about that decade which improved minority (but not white) scores. Most likely this was the general improvement in minorities' conditions around that time, giving them better nutrition and a more stable family life. It's hard to construct a narrative where it was school spending that did it – and even if it did, note that the majority of the increase in school spending happened from 1985 on, and demonstrably helped neither whites *nor* minorities.

I discuss this phenomenon more [here](#) and [here](#), but the summary is: no, it's not just because of special ed; no, it's not just a factor of how you measure test scores; no, there's not a "ceiling effect". Costs really did more-or-less double without any concomitant increase in measurable quality.

So, imagine you're a poor person. White, minority, whatever. Which would you prefer? Sending your child to a 2016 school? Or sending your child to a 1975 school, and getting a check for \$5,000 every year?

I'm proposing that choice because as far as I can tell that *is* the stakes here. 2016 schools have whatever tiny test score advantage they have over 1975 schools, and cost \$5000/year more, inflation adjusted. That \$5000 comes out of the pocket of somebody – either taxpayers, or other people who could be helped by government programs.

Second, college is even worse:



Note this is not adjusted for inflation; see link below for adjusted figures

Inflation-adjusted cost of a university education was [something like \\$2000/year in 1980](#). Now it's closer to \$20,000/year. No, it's not because of [decreased government funding](#), and there are similar trajectories for public and private schools.

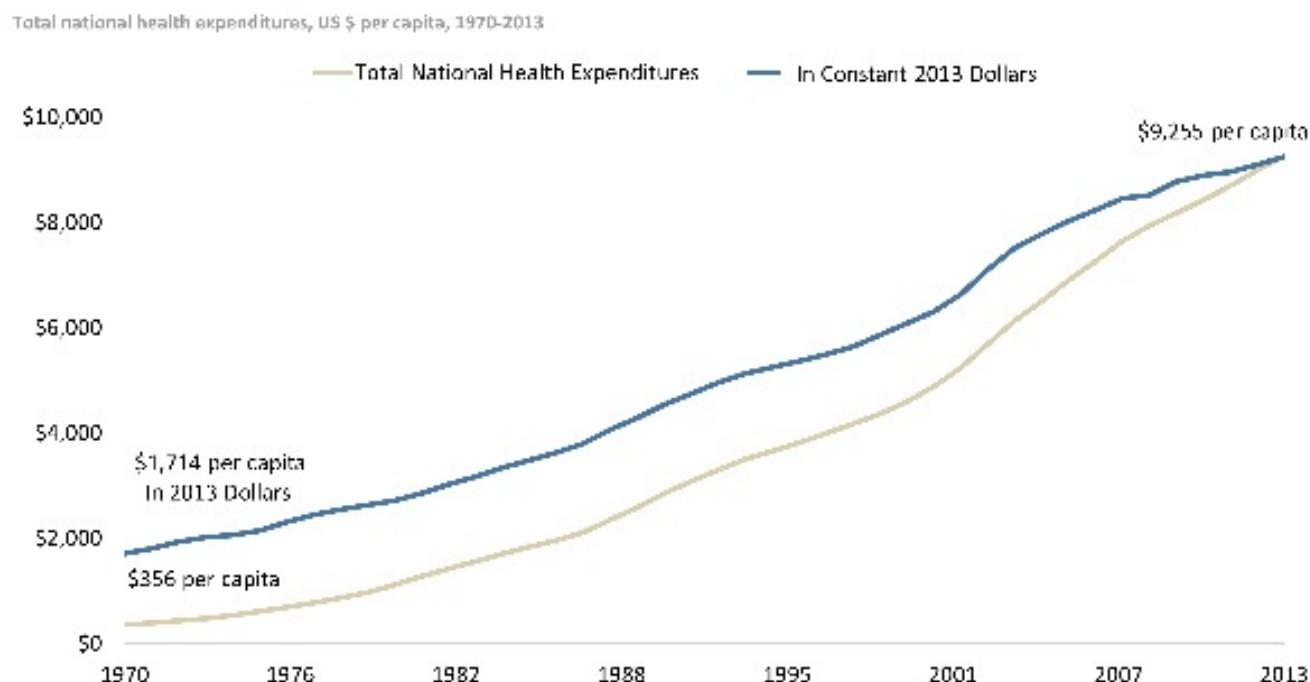
I don't know if there's an equivalent of "test scores" measuring how well colleges perform, so just use your best judgment. Do you think that modern colleges provide \$18,000/year greater value than colleges did in your parents' day? Would you rather graduate from a modern college, or graduate from a college more like the one your parents went to, plus get a check for \$72,000?

(or, more realistically, have \$72,000 less in student loans to pay off)

Was your parents' college even noticeably worse than yours? My parents sometimes talk about their college experience, and it seems to have had all the relevant features of a college experience. Clubs. Classes. Professors. Roommates. I might have gotten something extra for my \$72,000, but it's hard to see what it was.

Third, health care. The graph is starting to look disappointingly familiar:

On a per capita basis, health spending has grown substantially



Source: National Health Expenditure (NHE) data from Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group
Peterson-Kaiser Health System Tracker

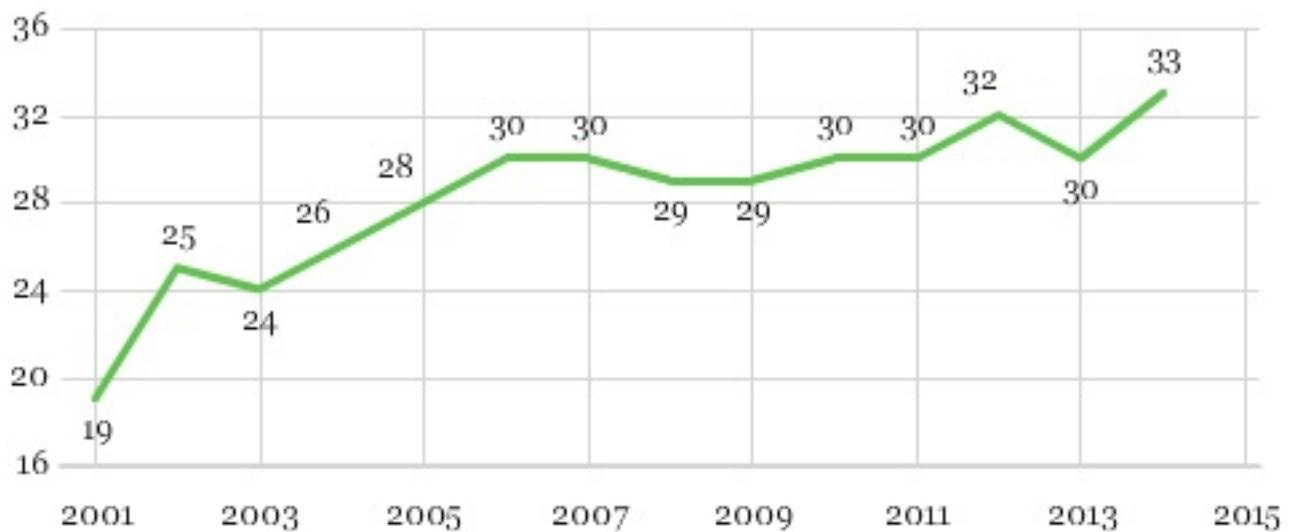
The cost of health care has about quintupled since 1970. It's actually been rising since earlier than that, but I can't find a good graph; it looks like it would have been about \$1200 in today's dollars in 1960, for an increase of about 800% in those fifty years.

This has had the expected effects. The average 1960 worker spent ten days' worth of their yearly paycheck on health insurance; the average modern worker spends sixty days' worth of it, a sixth of their entire earnings.

Percentage of Americans Putting Off Medical Treatment Because of Cost

Within the last 12 months, have you or a member of your family put off any sort of medical treatment because of the cost you would have to pay?

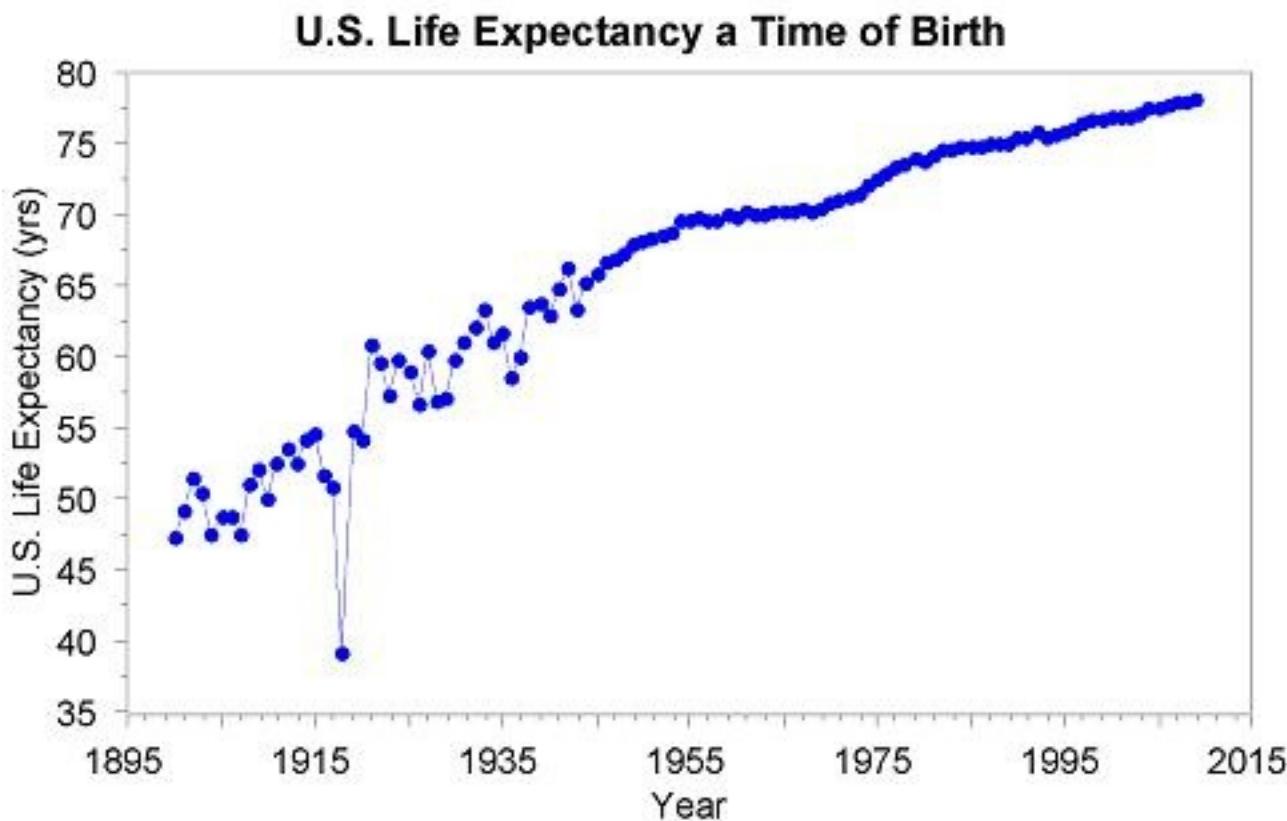
■ % Yes



GALLUP®

Or not.

This time I can't say with 100% certainty that all this extra spending has been for nothing. Life expectancy has gone way up since 1960:



Extra bonus conclusion: the Spanish flu was really bad

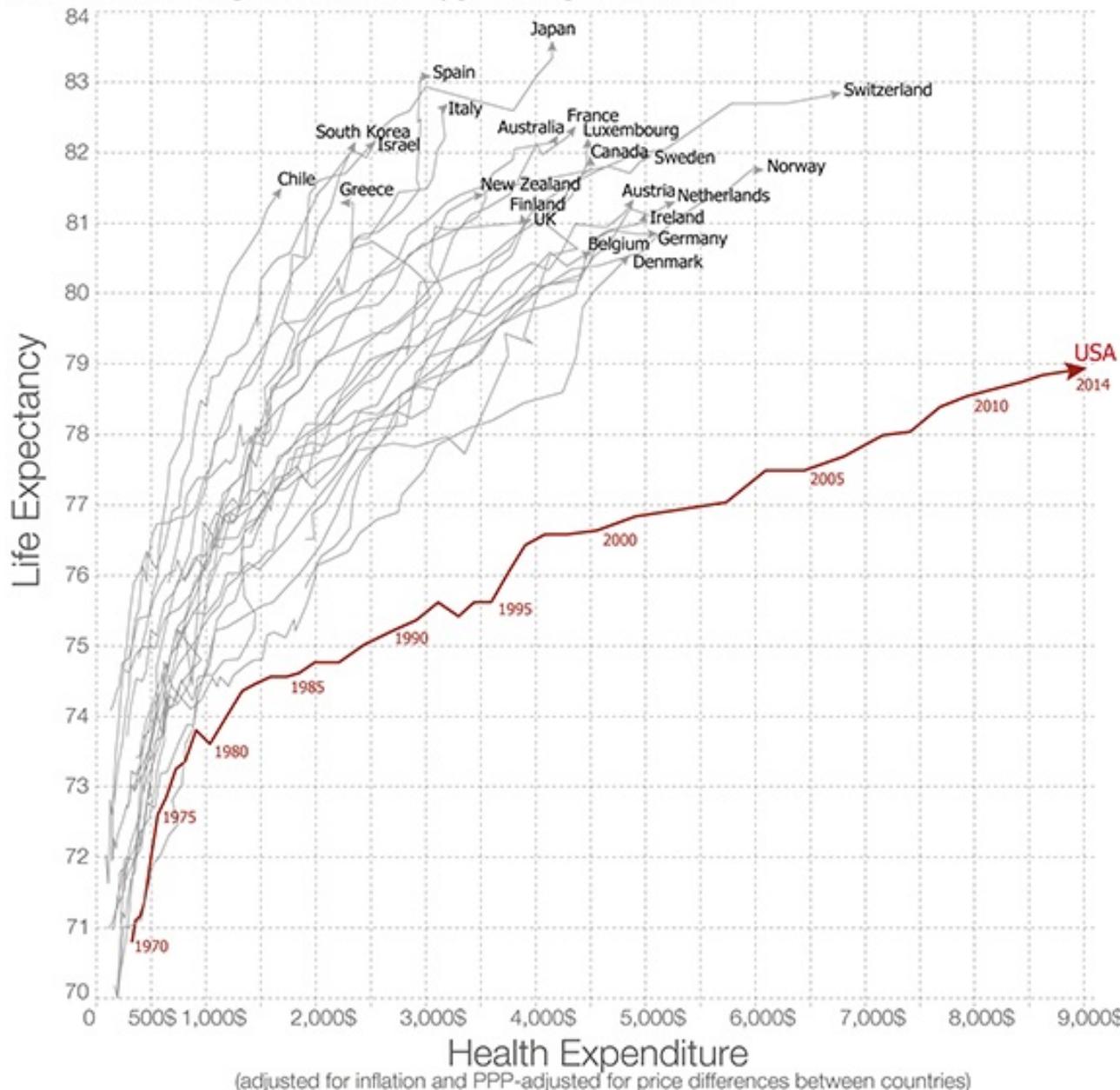
But a lot of people think that life expectancy depends on other things a lot more than healthcare spending. Sanitation, nutrition, quitting smoking, plus advances in health technology that don't involve spending more money. ACE inhibitors (invented in 1975) are great and probably increased lifespan a lot, but they cost \$20 for a year's supply and replaced older drugs that cost about the same amount.

In terms of calculating how much lifespan gain healthcare spending has produced, we have a couple of options. Start with by country:

Life expectancy vs. health expenditure over time (1970-2014)

Our World
in Data

Health spending measures the consumption of health care goods and services, including personal health care (curative care, rehabilitative care, long-term care, ancillary services and medical goods) and collective services (prevention and public health services as well as health administration), but excluding spending on investments. Shown is total health expenditure (financed by public and private sources).



Data source: Health expenditure from the OECD; Life expectancy from the World Bank. Licensed under CC-BY-SA by the author Max Roser. The data visualization is available at OurWorldInData.org and there you find more research and visualizations on this topic.

Countries like South Korea and Israel have about the same life expectancy as the US but pay about 25% of what we do. Some people use this to prove the superiority of centralized government health systems, although [Random Critical Analysis](#) has an alternative perspective. In any case, it seems very possible to get the same improving life expectancies as the US without octupling health care spending.

The Netherlands increased their health budget by a lot around 2000, sparking a bunch of studies on whether that increased life expectancy or not. There's a good meta-analysis [here](#), which lists six studies trying to calculate how much of the change in life expectancy was due to the large increases in health spending during this period. There's a

broad range of estimates: 0.3%, 1.8%, 8.0%, 17.2%, 22.1%, 27.5% (I'm taking their numbers for men; the numbers for women are pretty similar). They also mention two studies that they did not officially include; one finding 0% effect and one finding 50% effect (I'm not sure why these studies weren't included). They add:

In none of these studies is the issue of reverse causality addressed; sometimes it is not even mentioned. This implies that the effect of health care spending on mortality may be overestimated.

They say:

Based on our review of empirical studies, we conclude that it is likely that increased health care spending has contributed to the recent increase in life expectancy in the Netherlands. Applying the estimates from published studies to the observed increase in health care spending in the Netherlands between 2000 and 2010 [of 40%] would imply that 0.3% to almost 50% of the increase in life expectancy may have been caused by increasing health care spending. An important reason for the wide range in such estimates is that they all include methodological problems highlighted in this paper. However, this wide range indicates that the counterfactual study by Meerding et al, which argued that 50% of the increase in life expectancy in the Netherlands since the 1950s can be attributed to medical care, can probably be interpreted as an upper bound.

It's going to be completely irresponsible to try to apply this to the increase in health spending in the US over the past 50 years, since this is probably different at every margin and the US is not the Netherlands and the 1950s are not the 2010s. But if we irresponsibly take their median estimate and apply it to the current question, we get that increasing health spending in the US has been worth about one extra year of life expectancy. [This study](#) attempts to directly estimate a GDP corresponds to an increase of 0.05 years life expectancy. That would suggest a slightly different number of 0.65 years life expectancy gained by healthcare spending since 1960)

If these numbers seem absurdly low, remember [all of those controlled experiments](#) where giving people insurance doesn't seem to make them much healthier in any meaningful way.

Or instead of slogging through the statistics, we can just ask the same question as before. Do you think the average poor or middle-class person would rather:

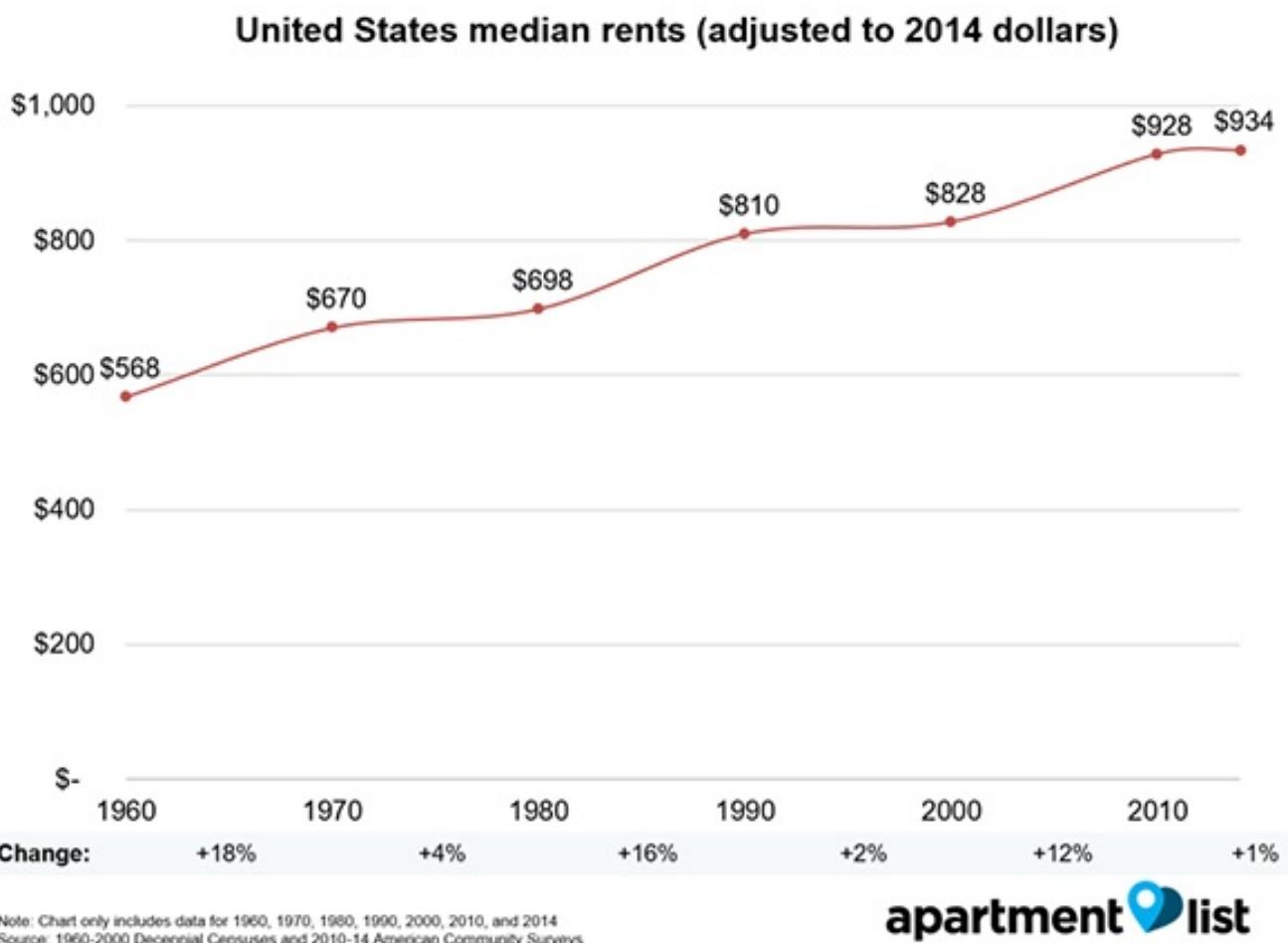
- a. Get modern health care
- b. Get the same amount of health care as their parents' generation, but with modern technology like ACE inhibitors, and also earn \$8000 extra a year

Fourth, we see similar effects in infrastructure. The first New York City subway opened around 1900. Various sources list lengths from 10 to 20 miles and costs from \$30 million to \$60 million dollars – I think my sources are capturing it at different stages of construction with different numbers of extensions. In any case, it suggests costs of between \$1.5 million to \$6 million dollars/mile = \$1-4 million per kilometer. That looks like it's about the inflation-adjusted equivalent of \$100 million/kilometer today, though I'm very uncertain about that estimate. In contrast, [Vox notes](#) that a new New York subway line being opened this year costs about *\$2.2 billion* per kilometer, suggesting a cost increase of twenty times – although I'm very uncertain about this estimate.

Things become clearer when you compare them country-by-country. The same Vox article notes that Paris, Berlin, and Copenhagen subways cost about \$250 million per kilometer, almost 90% less. Yet even those European subways are overpriced [compared to Korea](#), where a kilometer of subway in Seoul costs \$40 million/km (another Korean subway project cost \$80 million/km). This is a difference of 50x between Seoul and New York for apparently comparable

services. It suggests that the 1900s New York estimate above may have been roughly accurate if their efficiency was roughly in line with that of modern Europe and Korea.

Fifth, housing ([source](#)):



Most of the important commentary on this graph has [already been said](#), but I would add that optimistic takes like [this one](#) by the American Enterprise Institute are missing some of the dynamic. Yes, homes are bigger than they used to be, but part of that is zoning laws which make it easier to get big houses than small houses. There are a lot of people who would prefer to have a smaller house but don't. When I first moved to Michigan, I lived alone in a three bedroom house because there were no good one-bedroom houses available near my workplace and all of the apartments were loud and crime-y.

Or, once again, just ask yourself: do you think most poor and middle class people would rather:

1. Rent a modern house/apartment
2. Rent the sort of house/apartment their parents had, for half the cost

II

So, to summarize: in the past fifty years, education costs have doubled, college costs have decupled, health insurance costs have decupled, subway costs have at least decupled, and housing costs have increased by about fifty percent. US health care costs about four times as much as equivalent health care in other First World countries; US subways cost about eight times as much as equivalent subways in other First World countries.

I worry that people don't appreciate how weird this is. I didn't appreciate it for a long time. I guess I just figured that Grandpa used to talk about how back in his day movie tickets only cost a nickel; that was just the way of the world. *But all of the numbers above are inflation-adjusted.* These things have decupled in cost even *after* you adjust for movies costing a nickel in Grandpa's day. They have really, genuinely decupled in cost, no economic trickery involved.

And this is especially strange because we expect that improving technology and globalization ought to cut costs. In 1983, the first mobile phone cost \$4,000 – about \$10,000 in today's dollars. It was also a gigantic piece of crap. Today you can get a much better phone for \$100. This is the right and proper way of the universe. It's why we fund scientists, and pay businesspeople the big bucks.

But things like college and health care have *still* had their prices decuple. Patients can now schedule their appointments online; doctors can send prescriptions through the fax, pharmacies can keep track of medication histories on centralized computer systems that interface with the cloud, nurses get automatic reminders when they're giving two drugs with a potential interaction, insurance companies accept payment through credit cards – and all of this costs ten times as much as it did in the days of punch cards and secretaries who did calculations by hand.

It's actually even worse than this, because we take so many opportunities to save money that were unavailable in past generations. Underpaid foreign nurses immigrate to America and work for a song. Doctors' notes are sent to India overnight where they're transcribed by sweatshop-style labor for pennies an hour. Medical equipment gets manufactured in goodness-only-knows which obscure Third World country. And it *still* costs ten times as much as when this was all made in the USA – and that back when minimum wages were proportionally higher than today.

And it's actually even worse than *this*. A lot of these services have decreased in quality, presumably as an attempt to cut costs even further. Doctors used to make house calls; even when I was young in the '80s my father would still go to the houses of difficult patients who were too sick to come to his office. [This study](#) notes that for women who give birth in the hospital, "the standard length of stay was 8 to 14 days in the 1950s but declined to less than 2 days in the mid-1990s". The doctors I talk to say this isn't because modern women are healthier, it's because they kick them out as soon as it's safe to free up beds for the next person. Historic records of hospital care generally describe leisurely convalescence periods and making sure somebody felt absolutely well before letting them go; this seems bizarre to anyone who has participated in a modern hospital, where the mantra is to kick people out as soon as they're "stable" ie not in acute crisis.

If we had to provide the same quality of service as we did in 1960, and without the gains from modern technology and globalization, who even *knows* how many times more health care would cost? Fifty times more? A hundred times more?

And the same is true for colleges and houses and subways and so on.

III

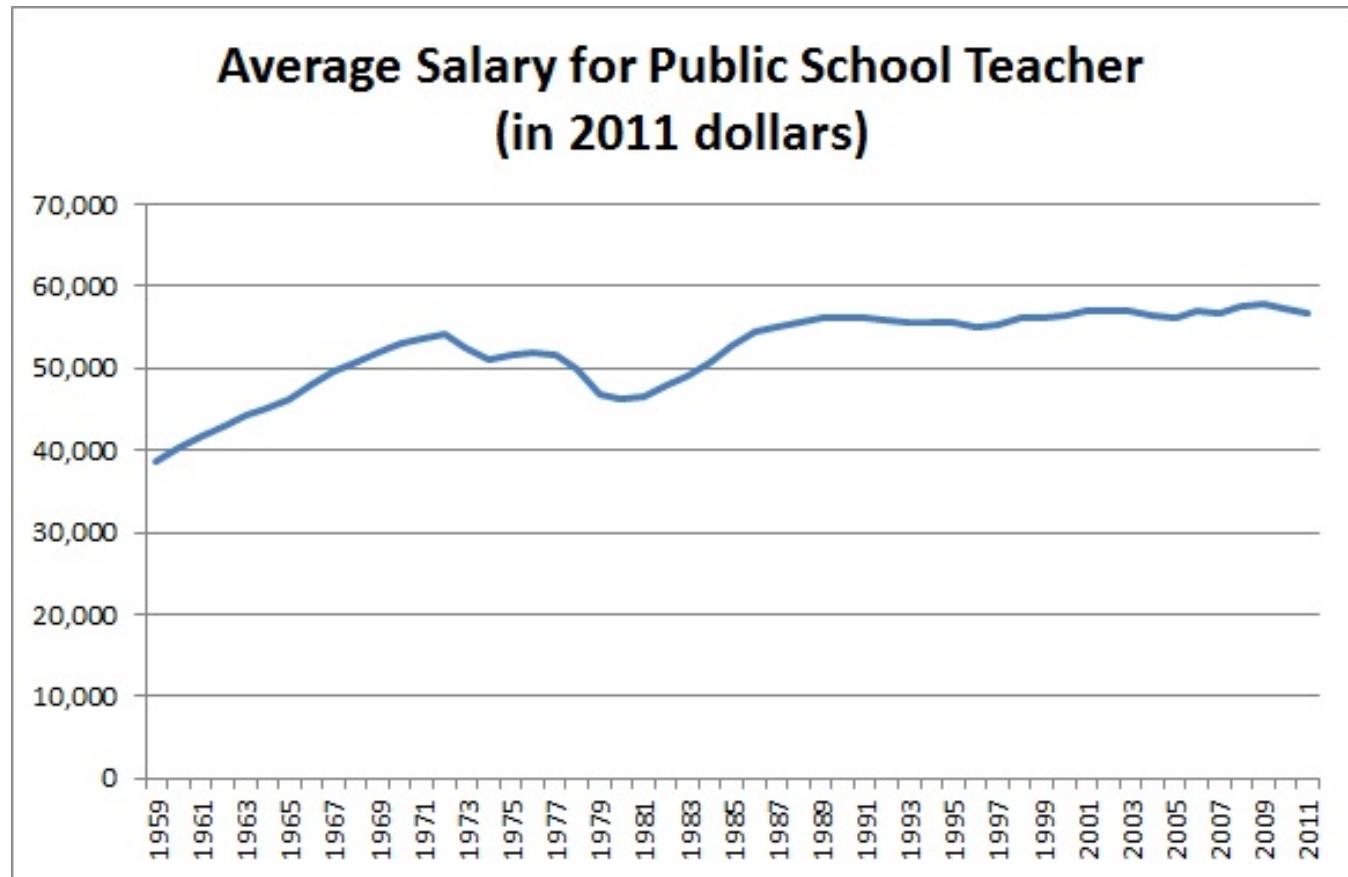
The existing literature on cost disease focuses on [the Baumol effect](#). Suppose in some underdeveloped economy, people can choose either to work in a factory or join an orchestra, and the salaries of factory workers and orchestra musicians reflect relative supply and demand and profit in those industries. Then the economy undergoes a technological revolution, and factories can produce ten times as many goods. Some of the increased productivity

trickles down to factory workers, and they earn more money. Would-be musicians leave the orchestras behind to go work in the higher-paying factories, and the orchestras have to raise their prices if they want to be assured enough musicians. So tech improvements in the factory sector raise prices in the orchestra sector.

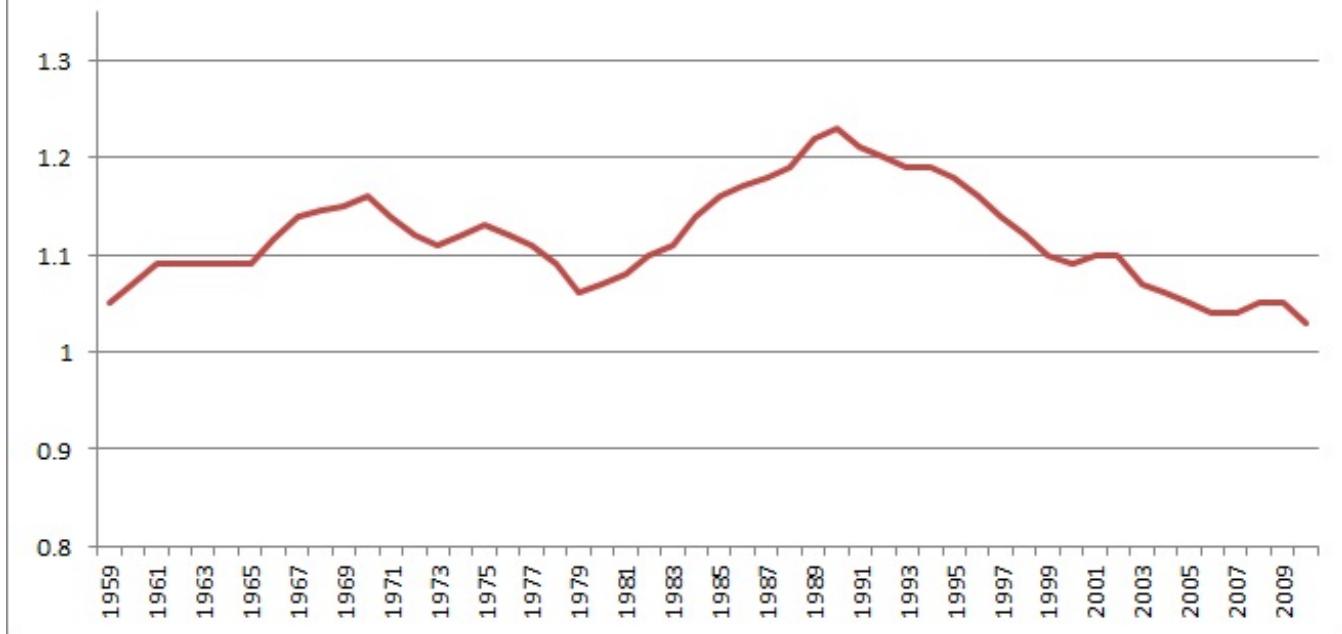
We could tell a story like this to explain rising costs in education, health care, etc. If technology increases productivity for skilled laborers in other industries, then less susceptible industries might end up footing the bill since they have to pay their workers more.

There's only one problem: health care and education aren't paying their workers more; in fact, quite the opposite.

Here are teacher salaries over time ([source](#)):

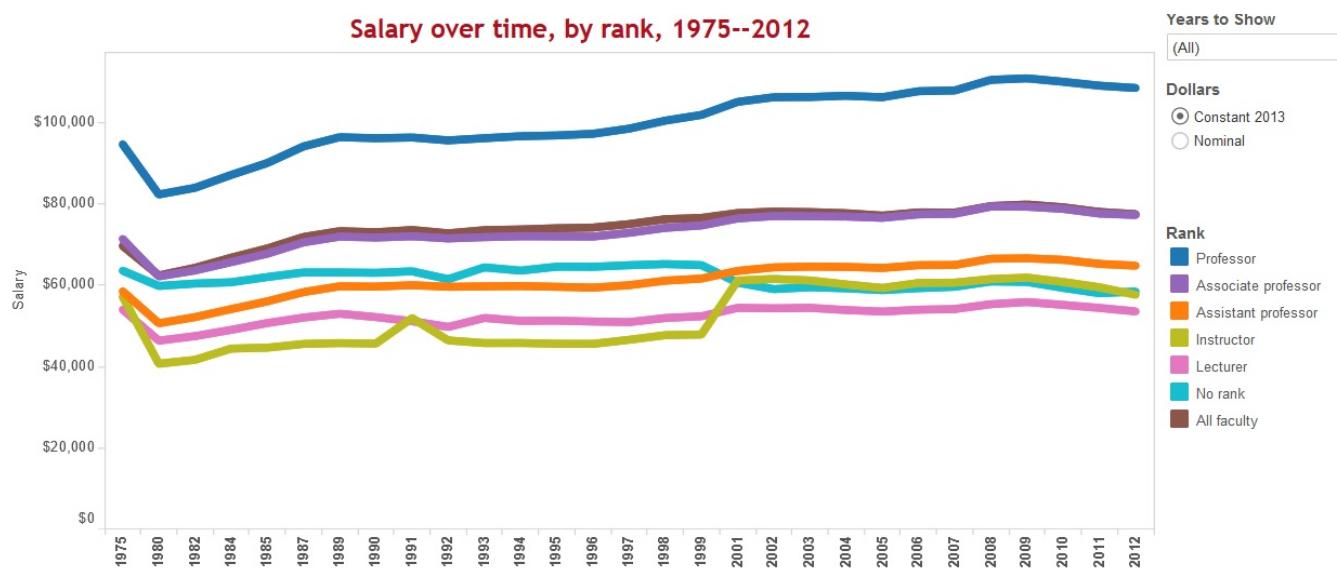


Average Salary for Public School Teacher, as Ratio fo Average Salary for All Full-Time Workers



Teacher salaries are relatively flat adjusting for inflation. But salaries for other jobs are increasing modestly relative to inflation. So teacher salaries relative to other occupations' salaries are actually declining.

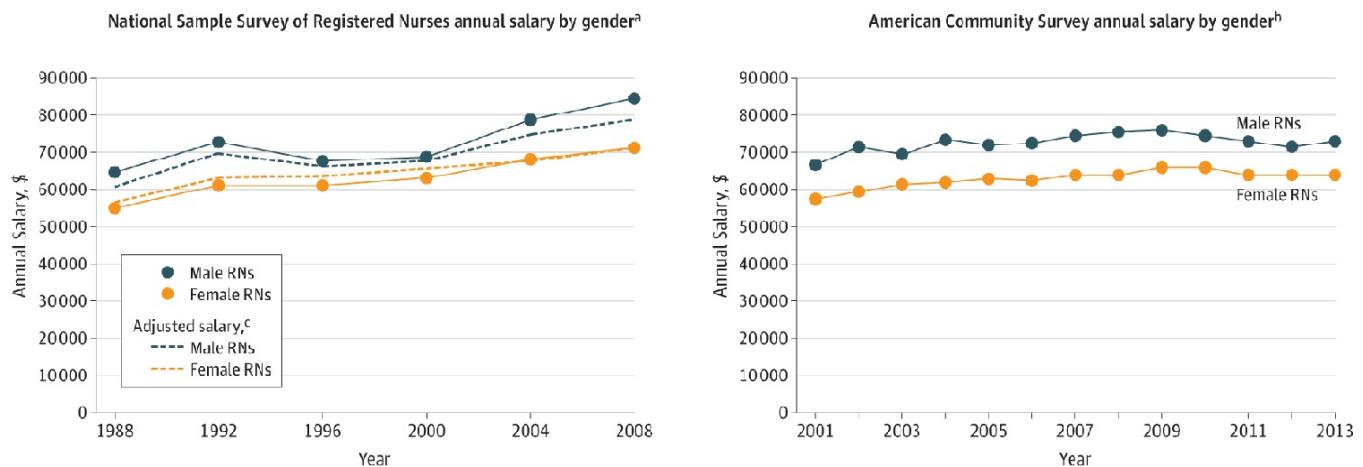
Here's a similar graph for professors ([source](#)):



Professor salaries are going up a little, but again, they're probably losing position relative to the average occupation. Also, note that although the average salary of each type of faculty is stable or increasing, the average salary of all faculty is going down. No mystery here – colleges are doing everything they can to switch from tenured professors to adjuncts, who complain of being [overworked and abused](#) while making about the same amount as a Starbucks barista.

This seems to me a lot like the case of the hospitals cutting care for new mothers. The price of the service dectuples, yet at the same time the service has to sacrifice quality in order to control costs.

And speaking of hospitals, here's the graph for nurses ([source](#)):



Female nurses' salaries went from about \$55,000 in 1988 to \$63,000 in 2013. This is probably around the average wage increase during that time. Also, some of this reflects [changes in education](#): in the 1980s only 40% of nurses had a degree; by 2010, about 80% did.

And for doctors ([source](#))



Stable again! Except that a lot of doctors' salaries now go to paying off their medical school debt, which has been ballooning like everything else.

I don't have a similar graph for subway workers, but come on. The overall pictures is that health care and education costs have managed to increase by ten times without a single cent of the gains going to teachers, doctors, or nurses. Indeed these professions seem to have lost ground salary-wise relative to others.

I also want to add some anecdote to these hard facts. My father is a doctor and my mother is a teacher, so I got to hear a lot about how these professions have changed over the past generation. It seems at least a little like the adjunct story, although without the clearly defined "professor vs. adjunct" dichotomy that makes it so easy to talk about. Doctors are really, really, *really* unhappy. When I went to medical school, some of my professors would tell me outright that they couldn't believe anyone would still go into medicine with all of the new stresses and demands placed on doctors. This doesn't seem to be limited to one medical school. *Wall Street Journal: Why Doctors Are Sick Of Their Profession* – "American physicians are increasingly unhappy with their once-vaunted profession, and that malaise is bad for their patients". *The Daily Beast: How Being A Doctor Became The Most Miserable Profession* – "Being a doctor has become a miserable and humiliating undertaking. Indeed, many doctors feel that America has declared war on physicians". *Forbes: Why Are Doctors So Unhappy?* – "Doctors have become like everyone else: insecure, discontent and scared about the future." *Vox: Only Six Percent Of Doctors Are Happy With Their Jobs*. *Al Jazeera America: Here's Why Nine Out Of Ten Doctors Wouldn't Recommend Medicine As A Profession*. Read these articles and they all say the same thing that all the doctors I know say – medicine used to be a well-respected,

enjoyable profession where you could give patients good care and feel self-actualized. Now it kind of sucks.

Meanwhile, I also see articles like [this piece from NPR](#) saying teachers are experiencing historic stress levels and up to 50% say their job “isn’t worth it”. Teacher job satisfaction is at [historic lows](#). And the veteran teachers I know say the same thing as the veteran doctors I know – their jobs used to be enjoyable and make them feel like they were making a difference; now they feel overworked, unappreciated, and trapped in mountains of paperwork.

It might make sense for these fields to become more expensive if their employees’ salaries were increasing. And it might make sense for salaries to stay the same if employees instead benefitted from lower workloads and better working conditions. But neither of these are happening.

IV

So what’s going on? Why *are* costs increasing so dramatically? Some possible answers:

First, can we dismiss all of this as an illusion? Maybe adjusting for inflation is harder than I think. Inflation is an average, so some things have to have higher-than-average inflation; maybe it’s education, health care, etc. Or maybe my sources have the wrong statistics.

But I don’t think this is true. The last time I talked about this problem, someone mentioned they’re running a private school which does just as well as public schools but costs only \$3000/student/year, a fourth of the usual rate.

Marginal Revolution [notes that](#) India has a private health system that delivers the same quality of care as its public system for a quarter of the cost. Whenever the same drug is provided by the official US health system and some kind of grey market supplement sort of thing, the grey market supplement costs between a fifth and a tenth as much; for example, Google’s first hit for Deplin®, official prescription L-methylfolate, [costs \\$175 for a month’s supply](#); unregulated L-methylfolate supplement delivers the same dose [for about \\$30](#). And this isn’t even mentioning things like [the \\$1 bag of saline that costs \\$700 at hospitals](#). Since it seems like it’s not too hard to do things for a fraction of what we currently do things for, probably we should be less reluctant to believe that the cost of everything is really inflated.

Second, might markets just not work? I know this is kind of an extreme question to ask in a post on economics, but maybe nobody knows what they’re doing in a lot of these fields and people can just increase costs and not suffer any decreased demand because of it. Suppose that people proved beyond a shadow of a doubt that Khan Academy could teach you just as much as a normal college education, but for free. People would still ask questions like – will employers accept my Khan Academy degree? Will it look good on a resume? Will people make fun of me for it? The same is true of community colleges, second-tier colleges, for-profit colleges, et cetera. I got offered a free scholarship to a mediocre state college, and I turned it down on the grounds that I knew nothing about anything and maybe years from now I would be locked out of some sort of Exciting Opportunity because my college wasn’t prestigious enough. Assuming everyone thinks like this, can colleges just charge whatever they want?

Likewise, my workplace offered me three different health insurance plans, and I chose the middle-expensiveness one, on the grounds that I had no idea how health insurance worked but maybe if I bought the cheap one I’d get sick and regret my choice, and maybe if I bought the expensive one I *wouldn’t* be sick and regret my choice. I am a doctor, my employer is a hospital, and the health insurance was for treatment in my own health system. The moral of the story is that I am an idiot. The second moral of the story is that people probably are not super-informed health care consumers.

This can’t be pure price-gouging, since corporate profits haven’t increased nearly enough to be where all the money is going. But a while ago a commenter linked me to the [Delta Cost Project](#), which scrutinizes the exact causes of

increasing college tuition. Some of it is the administrative bloat that you would expect. But a lot of it is fun “student life” types of activities like clubs, festivals, and paying Milo Yiannopoulos to speak and then cleaning up after the ensuing riots. These sorts of things improve the student experience, but I’m not sure that the average student would rather go to an expensive college with clubs/festivals/Milo than a cheap college without them. More important, it doesn’t really seem like the average student is offered this choice.

This kind of suggests a picture where colleges expect people will pay whatever price they set, so they set a very high price and then use the money for cool things and increasing their own prestige. Or maybe clubs/festivals/Milo become such a signal of prestige that students avoid colleges that don’t comply since they worry their degrees won’t be respected? Some people have pointed out that hospitals have switched from many-people-all-in-a-big-ward to private rooms. Once again, nobody seems to have been offered the choice between expensive hospitals with private rooms versus cheap hospitals with roommates. It’s almost as if industries have their own reasons for switching to more-bells-and-whistles services that people don’t necessarily want, and consumers just go along with it because for some reason they’re not exercising choice the same as they would in other markets.

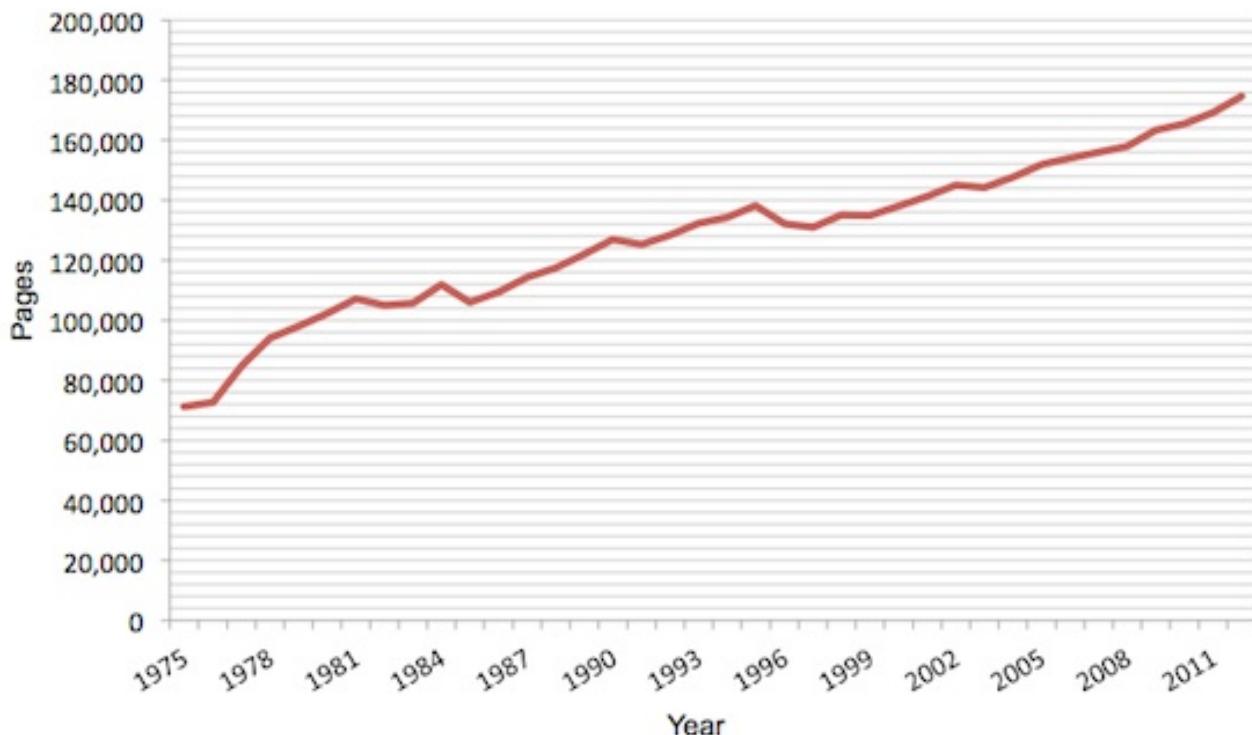
(this article on [the Oklahoma City Surgery Center](#) might be about a partial corrective for this kind of thing)

Third, can we attribute this to the inefficiency of government relative to private industry? I don’t think so. The government handles most primary education and subways, and has its hand in health care. But we know that for-profit hospitals aren’t much cheaper than government hospitals, and that private schools usually aren’t much cheaper (and are sometimes more expensive) than government schools. And private colleges cost more than government-funded ones.

Fourth, can we attribute it to indirect government intervention through regulation, which public and private companies alike must deal with? This seems to be at least part of the story in health care, given how much money you can save by grey-market practices that avoid the FDA. It’s harder to apply it to colleges, though some people have pointed out regulations like Title IX that affect the educational sector.

One factor that seems to speak out against this is that starting with Reagan in 1980, and picking up steam with Gingrich in 1994, we got an increasing presence of Republicans in government who declared war on overregulation – but the cost disease proceeded unabated. This is suspicious, but in fairness to the Republicans, they did sort of fail miserably at deregulating things. “The literal number of pages in the regulatory code” is kind of a blunt instrument, but it doesn’t exactly inspire confidence in the Republicans’ deregulation efforts:

Figure 1: Total Number of Pages in the Code of Federal Regulations 1975 -- 2011



Here's a more interesting (and more fun) argument against regulations being to blame: what about pet health care? Veterinary care is much less regulated than human health care, yet its cost is rising as fast (or faster) than that of the human medical system ([popular article](#), [study](#)). I'm not sure what to make of this.

Fifth, might the increased regulatory complexity happen not through literal regulations, but through fear of lawsuits? That is, might institutions add extra layers of administration and expense not because they're *forced* to, but because they fear being sued if they don't and then something goes wrong?

I see this all the time in medicine. A patient goes to the hospital with a heart attack. While he's recovering, he tells his doctor that he's really upset about all of this. Any normal person would say "You had a heart attack, of course you're upset, get over it." But if his doctor says this, and then a year later he commits suicide for some unrelated reason, his family can sue the doctor for "not picking up the warning signs" and win several million dollars. So now the doctor consults a psychiatrist, who does an hour-long evaluation, charges the insurance company \$500, and determines using her immense clinical expertise that the patient is upset because he just had a heart attack.

Those outside the field have *no idea* how much of medicine is built on this principle. [People often say that](#) the importance of lawsuits to medical cost increases is overrated because malpractice insurance doesn't cost that much, but the situation above would never look lawsuit-related; the whole thing only works because everyone involved documents it as well-justified psychiatric consult to investigate depression. Apparently [some studies](#) suggest this isn't happening, but all they do is survey doctors, and with all due respect all the doctors I know say the opposite.

This has nothing to do with government regulations (except insofar as these make lawsuits easier or harder), but it sure can drive cost increases, and it might apply to fields outside medicine as well.

Sixth, might we have changed our level of risk tolerance? That is, might increased caution be due not purely to lawsuitphobia, but to really caring more about whether or not people are protected? I read stuff every so often about how playgrounds are becoming obsolete because nobody wants to let kids run around unsupervised on something

with sharp edges. Suppose that one in 10,000 kids get a horrible playground-related injury. Is it worth making playgrounds cost twice as much and be half as fun in order to decrease that number to one in 100,000? This isn't a rhetorical question; I think different people can have legitimately different opinions here (though there are probably some utilitarian things we can do to improve them).

To bring back the lawsuit point, some of this probably relates to a difference between personal versus institutional risk tolerance. Every so often, an elderly person getting up to walk to the bathroom will fall and break their hip. This is a fact of life, and elderly people deal with it every day. Most elderly people I know don't spend thousands of dollars fall-proofing the route from their bed to their bathroom, or hiring people to watch them at every moment to make sure they don't fall, or buy a bedside commode to make bathroom-related falls impossible. This suggests a revealed preference that elderly people are willing to tolerate a certain fall probability in order to save money and convenience. Hospitals, which face huge lawsuits if any elderly person falls on the premises, are *not* willing to tolerate that probability. They put rails on elderly people's beds, place alarms on them that will go off if the elderly person tries to leave the bed without permission, and hire patient care assistants who among other things go around carefully holding elderly people upright as they walk to the bathroom ([I assume](#) this job will soon require at least a master's degree). As more things become institutionalized and the level of acceptable institutional risk tolerance becomes lower, this could shift the cost-risk tradeoff even if there isn't a population-level trend towards more risk-aversion.

Seventh, might things cost more for the people who pay because so many people don't pay? This is somewhat true of colleges, where an increasing number of people are getting in on scholarships funded by the tuition of non-scholarship students. I haven't been able to find great statistics on this, but one argument against: couldn't a college just not fund scholarships, and offer much lower prices to its paying students? I get that scholarships are good and altruistic, but it would be surprising if every single college thought of its role as an altruistic institution, and cared about it more than they cared about providing the same service at a better price. I guess this is related to my confusion about why more people don't open up colleges. Maybe this is the "smart people are rightly too scared and confused to go to for-profit colleges, and there's not enough ability to discriminate between the good and the bad ones to make it worthwhile to found a good one" thing again.

This also applies in health care. Our hospital (and every other hospital in the country) has some "frequent flier" patients who overdose on meth at least once a week. They come in, get treated for their meth overdose (we can't legally turn away emergency cases), get advised to get help for their meth addiction (without the slightest expectation that they will take our advice) and then get discharged. Most of them are poor and have no insurance, but each admission costs a couple of thousand dollars. The cost gets paid by a combination of taxpayers and other hospital patients with good insurance who get big markups on their own bills.

Eighth, might total compensation be increasing even though wages aren't? There definitely seems to be a pensions crisis, especially in a lot of government work, and it's possible that some of this is going to pay the pensions of teachers, etc. My understanding is that in general [pensions aren't really increasing much faster than wages](#), but this might not be true in those specific industries. Also, this might pass the buck to the question of why we need to spend more on pensions now than in the past. I don't think increasing life expectancy explains all of this, but I might be wrong.

IV

I mentioned politics briefly above, but they probably deserve more space here. Libertarian-minded people keep talking about how there's too much red tape and the economy is being throttled. And less libertarian-minded people keep interpreting it as not caring about the poor, or not understanding that government has an important role in a civilized society, or as a "dog whistle" for racism, or whatever. I don't know why more people don't just come out and say "LOOK, REALLY OUR MAIN PROBLEM IS THAT ALL THE MOST IMPORTANT THINGS COST TEN TIMES AS MUCH

AS THEY USED TO FOR NO REASON, PLUS THEY SEEM TO BE GOING DOWN IN QUALITY, AND NOBODY KNOWS WHY, AND WE'RE MOSTLY JUST DESPERATELY FLAILING AROUND LOOKING FOR SOLUTIONS HERE." State that clearly, and a lot of political debates take on a different light.

For example: some people promote free universal college education, remembering a time when it was easy for middle class people to afford college if they wanted it. Other people oppose the policy, remembering a time when people didn't depend on government handouts. Both are true! My uncle paid for his tuition at a really good college just by working a pretty easy summer job – not so hard when college cost a tenth of what it did now. The modern conflict between opponents and proponents of free college education is over how to distribute our losses. In the old days, we could combine low taxes with widely available education. Now we can't, and we have to argue about which value to sacrifice.

Or: some people get upset about teachers' unions, saying they must be sucking the "dynamism" out of education because of increasing costs. Others people fiercely defend them, saying teachers are underpaid and overworked. Once again, in the context of cost disease, both are obviously true. The taxpayers are just trying to protect their right to get education as cheaply as they used to. The teachers are trying to protect their right to make as much money as they used to. The conflict between the taxpayers and the teachers' unions is about how to distribute losses; *somebody* is going to have to be worse off than they were a generation ago, so who should it be?

And the same is true to greater or lesser degrees in the various debates over health care, public housing, et cetera.

Imagine if tomorrow, the price of water decupled. Suddenly people have to choose between drinking and washing dishes. Activists argue that taking a shower is a basic human right, and grumpy talk show hosts point out that in *their* day, parents taught their children not to waste water. A coalition promotes laws ensuring government-subsidized free water for poor families; a Fox News investigative report shows that some people receiving water on the government dime are taking long luxurious showers. Everyone gets really angry and there's lots of talk about basic compassion and personal responsibility and whatever but all of this is secondary to *why does water cost ten times what it used to?* I think this is the basic intuition behind so many people, even those who genuinely want to help the poor, are afraid of "tax and spend" policies. In the context of cost disease, these look like industries constantly doubling, tripling, or decupling their price, and the government saying "Okay, fine," and increasing taxes however much it costs to pay for whatever they're demanding now.

If we give everyone free college education, that solves a big social problem. It also locks in a price which is ten times too high for no reason. This isn't fair to the government, which has to pay ten times more than it should. It's not fair to the poor people, who have to face the stigma of accepting handouts for something they could easily have afforded themselves if it was at its proper price. And it's not fair to future generations if colleges take this opportunity to increase the cost by *twenty* times, and then our children have to subsidize *that*.

I'm not sure how many people currently opposed to paying for free health care, or free college, or whatever, would be happy to pay for health care that cost less, that was less wasteful and more efficient, and whose price we expected to go down rather than up with every passing year. I expect it would be a lot.

And if it isn't, who cares? The people who want to help the poor have enough political capital to spend eg \$500 billion on Medicaid; if that were to go ten times further, then everyone could get the health care they need without any more political action needed. If some government program found a way to give poor people good health insurance for a few hundred dollars a year, college tuition for about a thousand, and housing for only two-thirds what it costs now, that would be the greatest anti-poverty advance in history. That program is called "having things be as efficient as they were a few decades ago".

V

In 1930, economist John Maynard Keynes [predicted](#) that his grandchildrens' generation would have a 15 hour work week. At the time, it made sense. GDP was rising so quickly that anyone who could draw a line on a graph could tell that our generation would be four or five times richer than his. And the average middle-class person in his generation felt like they were doing pretty well and had most of what they needed. Why *wouldn't* they decide to take some time off and settle for a lifestyle merely twice as luxurious as Keynes' own?

Keynes was sort of right. GDP per capita *is* 4-5x greater today than in his time. Yet we still work forty hour weeks, and some large-but-inconsistently-reported percent of Americans ([76?](#) [55?](#) [47?](#)) still live paycheck to paycheck.

And yes, part of this is because [inequality is increasing and most of the gains are going to the rich](#). But this alone wouldn't be a disaster; we'd get to Keynes' utopia a little slower than we might otherwise, but eventually we'd get there. Most gains going to the rich means at least some gains are going to the poor. And at least there's a lot of mainstream awareness of the problem.

I'm more worried about the part where the cost of basic human needs goes up faster than wages do. Even if you're making twice as much money, if your health care and education and so on cost ten times as much, you're going to start falling behind. Right now the standard of living isn't just stagnant, it's [at risk of declining](#), and a lot of that is student loans and health insurance costs and so on.

What's happening? I don't know and I find it really scary.

What Happened To '90s Environmentalism?

Posted on January 1, 2019 by Scott Alexander



0. Introduction

I grew up in the 90s, which meant watching movies about plucky children fighting Pollution Demons. Sometimes teachers would show them to us in class. None of us found that strange. We knew that when we grew up, this would be our fight: to take on the loggers and whalers and seal-clubbers who were destroying our planet and save the Earth for the next generation.

What *happened* to that? I don't mean the Pollution Demons: they're still around, I think one of them runs Trump's EPA now. What happened to everything else? To those teachers, those movies, that whole worldview?

Save The Whales. Save The Rainforest. Save Endangered Species. Save The Earth. Stop Slash-And-Burn. Stop Acid Rain. Earth Day Every Day. Reduce, Reuse, Recycle. Twenty-five years ago, each of those would invoke a whole acrimonious debate; to some, a battle-cry; to others, a sign of a dangerous fanaticism that would destroy the economy. Today they sound about as relevant as "Fifty-four forty or fight" and "Remember the Maine". Old slogans, emptied of their punch and fit only for bloodless historical study.

If you went back in time, turned off our Pollution Demon movie, and asked us to predict what would come of the environment twenty-five years, later, in 2018, I think we would imagine one of two scenarios. In the first, the world had become a renewable ecotopia where every child was taught to live in harmony with nature. In the second, we had failed in our struggle, the skies were grey, the rivers were brown, wild animals were a distant memory – but at least a few plucky children would still be telling us it wasn't too late, that we could start the tough job of cleaning up after ourselves and changing paths to that other option.

The idea that things wouldn't really change – that the environment would neither move noticeably forward or noticeably backwards – but that everyone would stop talking about environmentalism – that you could go years without hearing the words "endangered species" – that nobody would even know whether the rainforests were expanding or contracting – wouldn't even be on the radar. It would sound like some kind of weird bizarro-world.

Just to prove I'm not imagining all this:

● rainforests

Search term

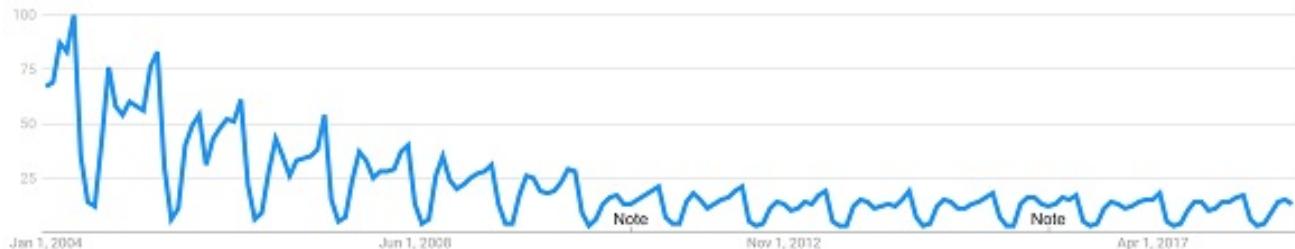
United States ▾

2004 - present ▾

All categories ▾

Web Search ▾

Interest over time ?



This is the volume of Google searches for “rainforests” over time. It goes up each year [when school starts](#), and crashes again for summer vacation. But on average, there are only about 18% as many rainforest-related searches today as in 2004.

● "endangered species"

Search term

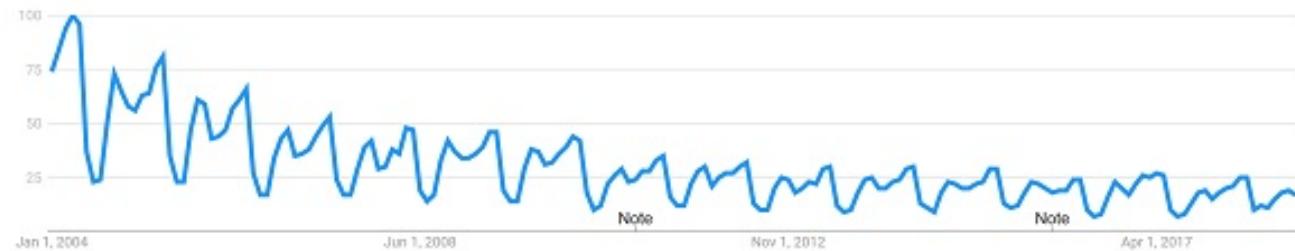


United States ▾

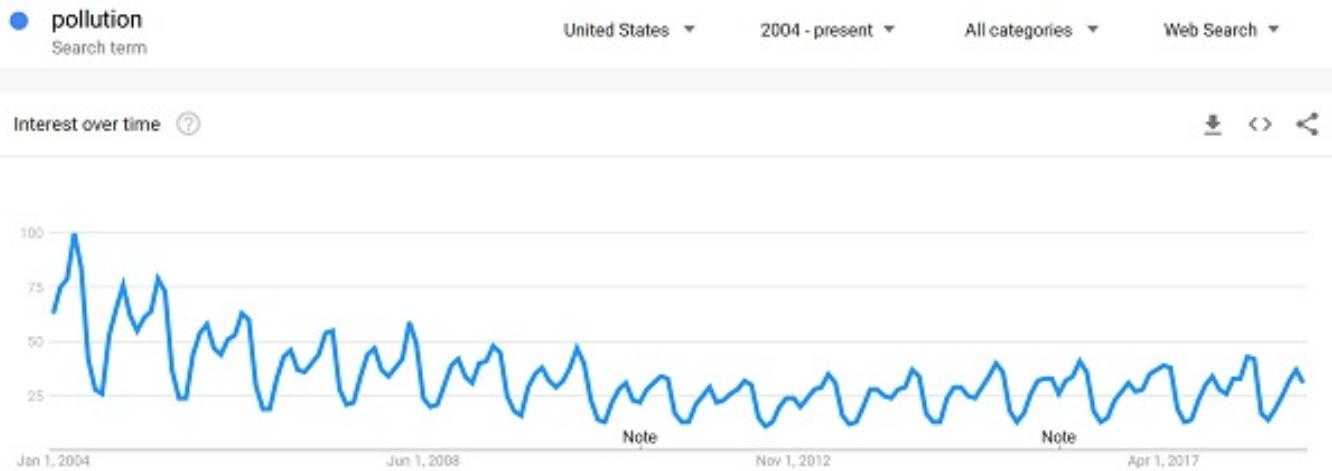
2004 - present ▾

All categories ▾

Interest over time ?



“Endangered species”, 25%



"Pollution", 43%

And these are just since Google started tracking searches in 2004. The decline of 90s environmentalism must be even bigger.

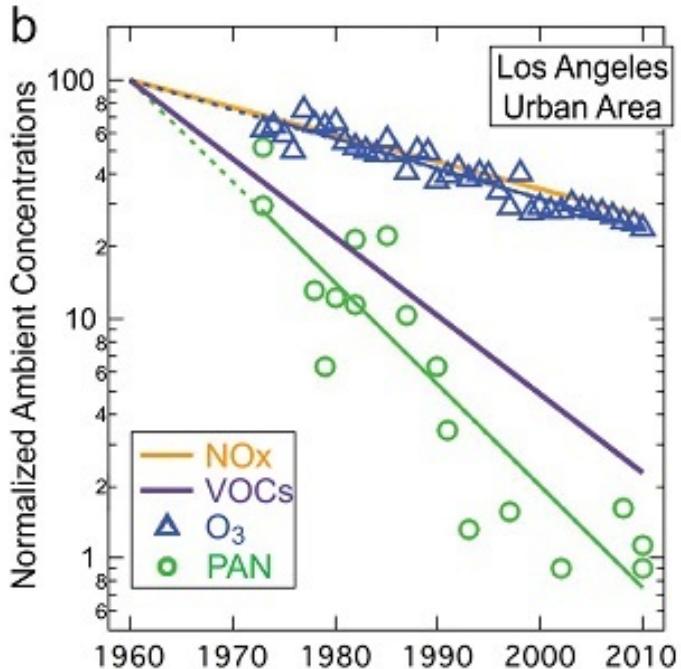
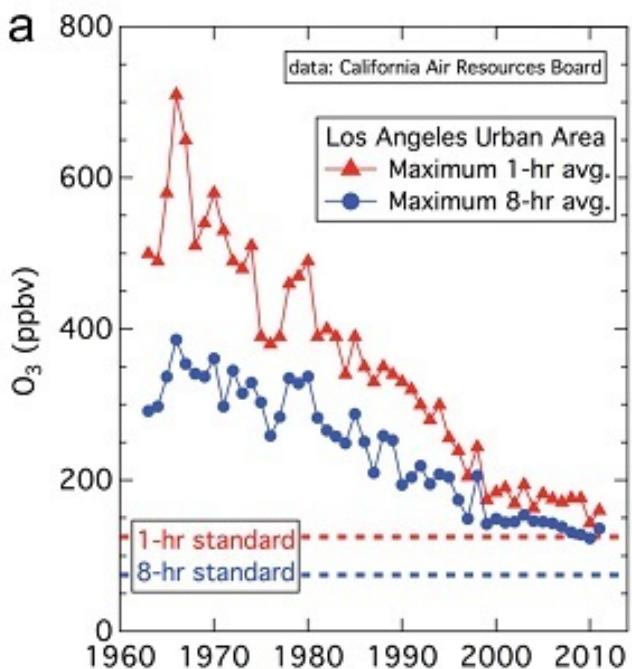
So *what happened?* Every so often you'll hear someone mutter darkly "You never hear about the ozone hole these days, guess that was a big nothingburger." This summons a horde of environmentalists competing to point out that you never hear about the ozone hole these days because environmentalists successfully fixed it. There was a [big conference](#) in 1989 where all the nations of the world met together and agreed to stop using ozone-destroying chlorofluorocarbons, and the ozone hole is recovering according to schedule. When people use the ozone hole as an argument against alarmism, environmentalism is a victim of its own success.

So what about these other issues that have since fizzled out? Did environmentalists solve them? Did they never exist in the first place? Or are they still as bad as ever, and we've just stopped caring?

1. Air And Water Pollution

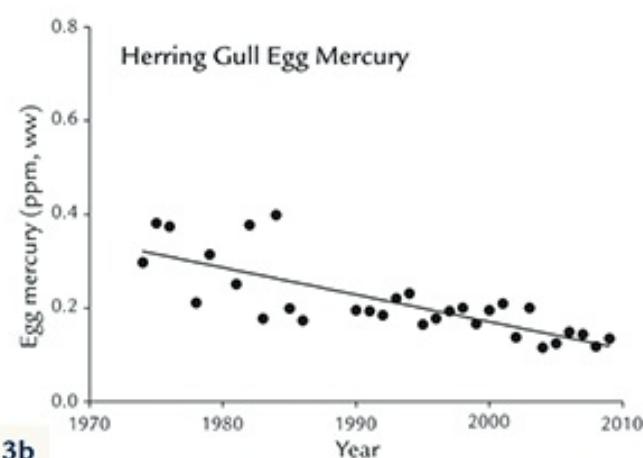
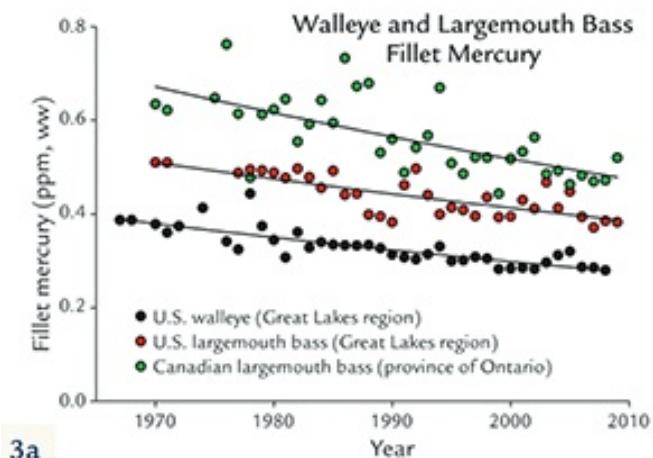
Have you seen what Chinese cities look like on a smoggy day? Trick question: neither have the Chinese. The US used to be like that. I grew up near Los Angeles during the 1990s. My mother tells the story of a time when I was very young and my grandparents came to visit from the Midwest. "It reminds us of home," they said, "it's so flat." "We're surrounded by mountains", my mother told them. We were. You couldn't see any of them.

Environmentalists crusaded against this. Here are the results:



A lot of the credit goes to the [Clean Air Act](#), passed in 1963 and tightened in 1990. Along with its more visible (pun intended) effects, scientists suspect it has prevented about 200,000 deaths from lung disease and a host of other cases of asthma, bronchitis, and even heart attacks.

It's hard to find great data on water because there are so many different kinds of water and so many different ways it can be polluted. But just to choose a random very bad thing, here's mercury levels in Great Lakes fish:



I don't know of anyone claiming this is anything other than a response to stricter environmental laws.

As a result of these victories, people are no longer as concerned about air and water pollution. From [Gallup](#):

This seems like a clear case of good work.

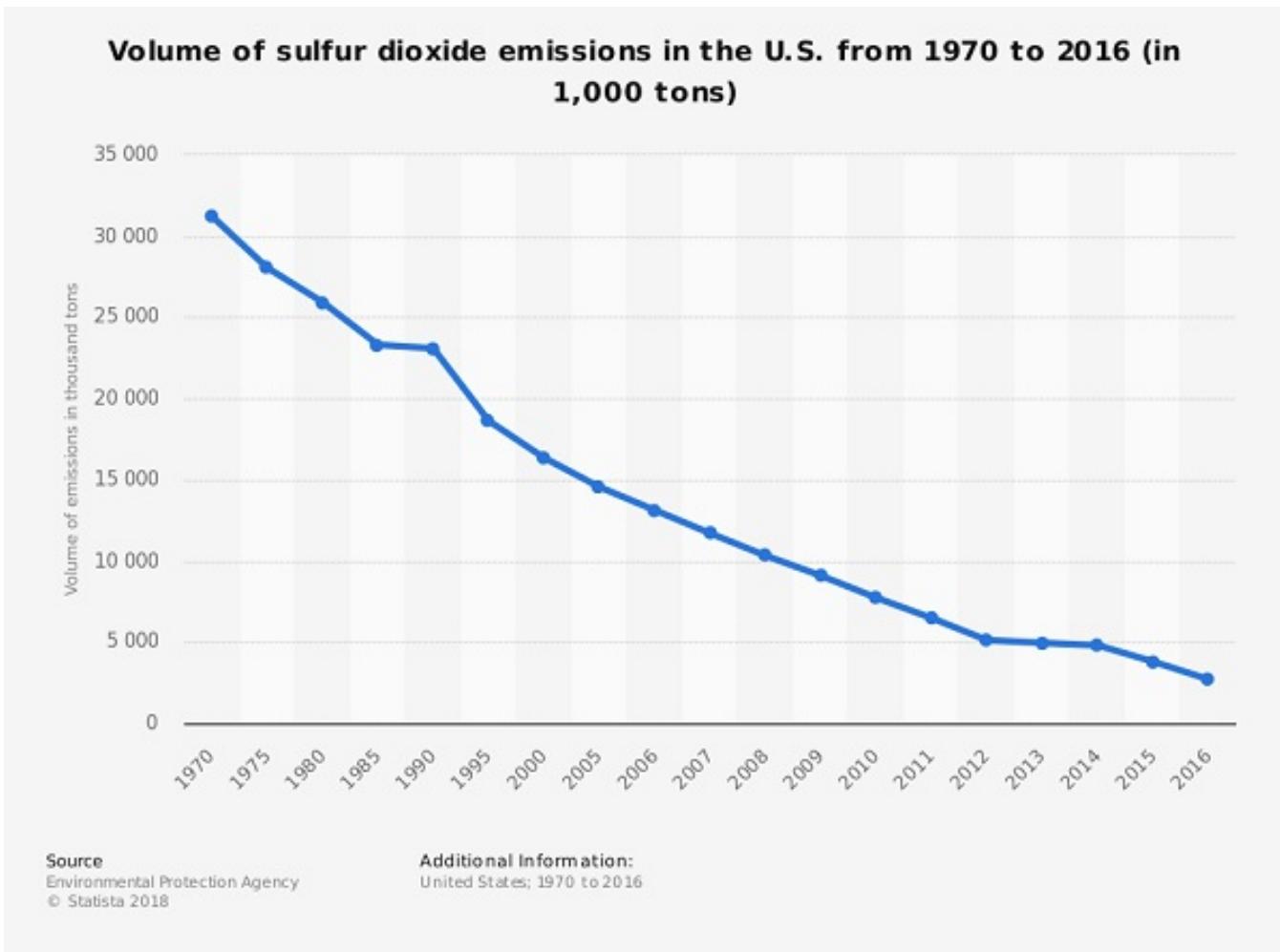
Verdict: Environmental movement successfully solved this problem.

2. Acid Rain

Acid rain is a combination of rain and pollution which gets very acidic and destroys plants and structures. It was a staple of very early 90s environmentalism, and understandably so: the prospect of acid falling from the sky and dissolving everything is very attention-grabbing. I remember the discourse focusing on statues; George Washington's marble face slowly melting under sizzling raindrops makes a heck of an image.

I am not the first person to notice that Washington's face remains mercifully unmelted. In 2009, Slate asked [Whatever Happened To Acid Rain?](#). EPA Blog, 2010: [Whatever Happened To Acid Rain?](#). 2012, Mental Floss: [What Ever Happened To Acid Rain?](#) By 2018 the Internet had advanced, so here's the [Whatever Happened To Acid Rain Podcast](#). Even the Encyclopedia Britannica, itself a good candidate for a "Whatever Happened To..." piece, has a [What Happened To Acid Rain](#) article.

Most of these sources say environmentalists solved acid rain by cutting down on emission of sulfur dioxide, the main offending chemical. A Bush I era cap-and-trade policy gets a lot of the credit in the US, but it looks like it was a broader effort than that:



There's less clear data on rain acidity, but all my sources agree it has modestly declined in the US, though it is still "between 2.5 and eight times more acidic than it should be". Lakes and rivers are slowly recovering. On the other hand, in newly-industrializing countries like China and India, rain is becoming more acidic and they're going through some of the same issues we were in the 80s.

This picture is slightly complicated by some people who claim acid rain was always exaggerated and "we solved it" is a convenient retreat from acknowledging this (for what it's worth, these people tend to be global warming skeptics too). Most of them point to the 1990 [National Acid Precipitation Assessment Program](#), a giant government investigation into the acid rain problem. I found a 1990 *New York Times* article on the report [here](#):

A comprehensive Federal report that was supposed to resolve the issue of how much damage is caused to forests by acid rain has come under criticism from some distinguished scientists who are reviewing it.

The critics said that the report gave an incorrect impression that air pollution was not causing any large-scale problems for forest ecosystems. They also said that the report, still in draft form, ignored a number of studies suggesting serious air pollution problems.

But other experts contend that the general conclusion of the report is essentially right. The report concluded that with the exception of damage to red spruce at high elevations in the East, forests in the United States are not suffering serious damage from acid rain [...]

The report now being reviewed is the final draft, completed at a cost of nearly \$500 million. It examines the

effects of other pollutants, like ozone, as well as acid deposits, and it concludes that air pollution causes far less environmental damage than has been feared.

An interim report issued by the study group in 1987, before Dr. Mahoney became director, was sharply criticized by many scientists. They contended that it tailored research findings into conclusions that matched the political goals of the Reagan Administration, which opposed new controls on air pollution. No such criticism has been leveled at the 28-volume final draft, which has been generally praised as a sound scientific document.

There is, however, some unhappiness among scientists with the volume dealing with forest health and productivity in the United States and Canada.

Dr. Ellis B. Cowling, associate for research at North Carolina State University's College of Forest Resources, said in a telephone interview: "The tone is that we don't have a problem except in southern California, and with red spruce at high altitudes. That is not a fair statement of the state of scientific knowledge." He added, "Perhaps the authors were a bit too hasty in reaching conclusions."

Dr. Cowling, who is highly regarded by colleagues as a conservative, solid scientist, wrote a memorandum to the authors of the forest health volume. He offered a series of suggestions for changing the wording of conclusions in ways that he said would reflect the state of science more accurately.

The first of those would change a finding that stated, "The vast majority of forests in the United States and Canada are not affected by decline." To be more consistent with the data, Dr. Cowling said, the conclusion should read: "Most forests in the United States do not show unusual visible symptoms of stress, marked decreases in the rate of growth or significant increases in mortality."

Just because symptoms of forest decline are not currently visible, Dr. Cowling argued, does not rule out the possibility that they are under way.

[This article](#) also provides a summary of contemporaneous responses to NAPAP, which quotes study director James Mahoney's summary of his own report: "The sky is not falling, but there is a problem that needs addressing."

I cannot find anyone really challenging the NAPAP report nowadays, so I provisionally accept that the damage from acid rain, while real, was exaggerated at the time.

There's a related debate about how much the lakes and streams affected have recovered. Some lakes and streams are naturally acidic; there is some debate over what percent of lake/stream acidity is natural vs. acid-rain-related. In recent years this debate has focused on whether lakes/streams have recovered after the SO₂ decline; if they haven't, this might suggest their problems were never human-activity-related in the first place.

Global warming skeptic blog Watt's Up With That [claims](#) they haven't:

Possibly the greatest evidence against harmful effects of acid rain is the fact that acidic lakes have not "recovered" after most sulfur and nitrogen pollution was removed from the atmosphere. The 2011 NAPAP report to Congress stated that SO₂ and NO₂ emissions were down, that airborne concentrations were down, and that acid deposition from rainfall was down, but could not report that lake acidity was significantly reduced. The report states, "Scientists have observed delays in ecosystem recovery in the eastern United States despite decreases in emissions and deposition over the last 30 years." In other words, the pollution was mostly eliminated, but the lakes are still acidic.

You can find [the report](#) here. Like all long government reports, the details are ten zillion different trends in different directions that don't form a cohesive narrative, and the executive summary is "things are good in all the ways that suggest we deserve more money, but bad in all the ways that suggest we need more money". It is complicated enough that you shouldn't trust my excerpting, but at least to me the relevant excerpts seem to be:

Levels of acid neutralizing capacity (ANC), an indicator of the ability of a waterbody to neutralize acid deposition, have shown improvement from 1990 to 2008 at many lake and stream long-term monitoring sites in the eastern United States, including New England and the Adirondack Mountains. Many lakes and streams still have acidic conditions harmful to their biota even though the increases in ANC indicate that some recovery from acidification is occurring in sensitive aquatic ecosystems

And:

Despite the environmental improvements reported here, research over the past few years indicates that recovery from the effects of acidification is not likely for many sensitive areas without additional decreases in acid deposition. Many published articles, as well as the modeling presented in this report, show that the SO₂ and NO_x emission reductions achieved under Title IV from power plants are not recognized as sufficient to achieve full recovery or to prevent further acidification in some regions.

So Watts seems to be mostly wrong when they say lakes are not recovering, but mostly right when they say ecosystems are not recovering. But NAPAP has some explanations for why ecosystems are not recovering: first, if you poison a lake and kill everything, then even if you remove the poison later everything is still dead. Second, there are complicated natural cycles that gradually wash old deposited land-based pollution into lakes, and it will be a long time before all the pollution deposited on land gets fully washed away. Third, maybe we haven't fought acid rain hard enough.

I think a lot of the epistemic work here is going to get done by people's respective stereotypes about the trustworthiness of global warming denialists vs. big government agencies whose budget depends on there being a problem. But my impression is that Watts' claim that poor recovery suggests acid rain was never a problem don't hold up very well.

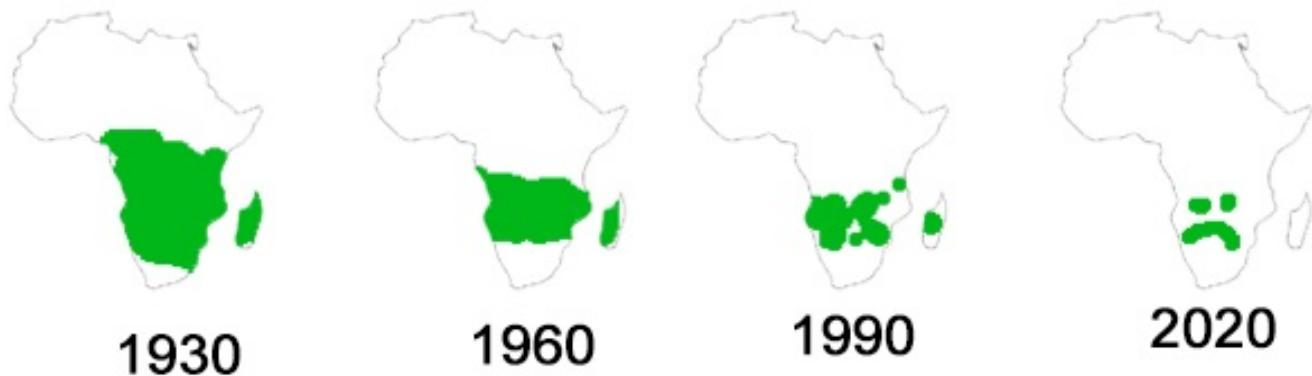
In any case, it's undeniable that rain has become a lot less acid lately, and likely that this has at least modest positive effects on some ecosystems as well as on the built environment. Anti-Confederate protesters have replaced acid rain as the number one threat to our statues. Our precious, precious statues. Someday they will be safe.

Verdict: *A little of everything: partly solved, partly alarmism, partly still going on.*

3. The Rainforests

Maybe the most typical image of 90s environmentalism is men in bulldozers clear-cutting a rainforest, while tapirs and tree sloths gently weep.

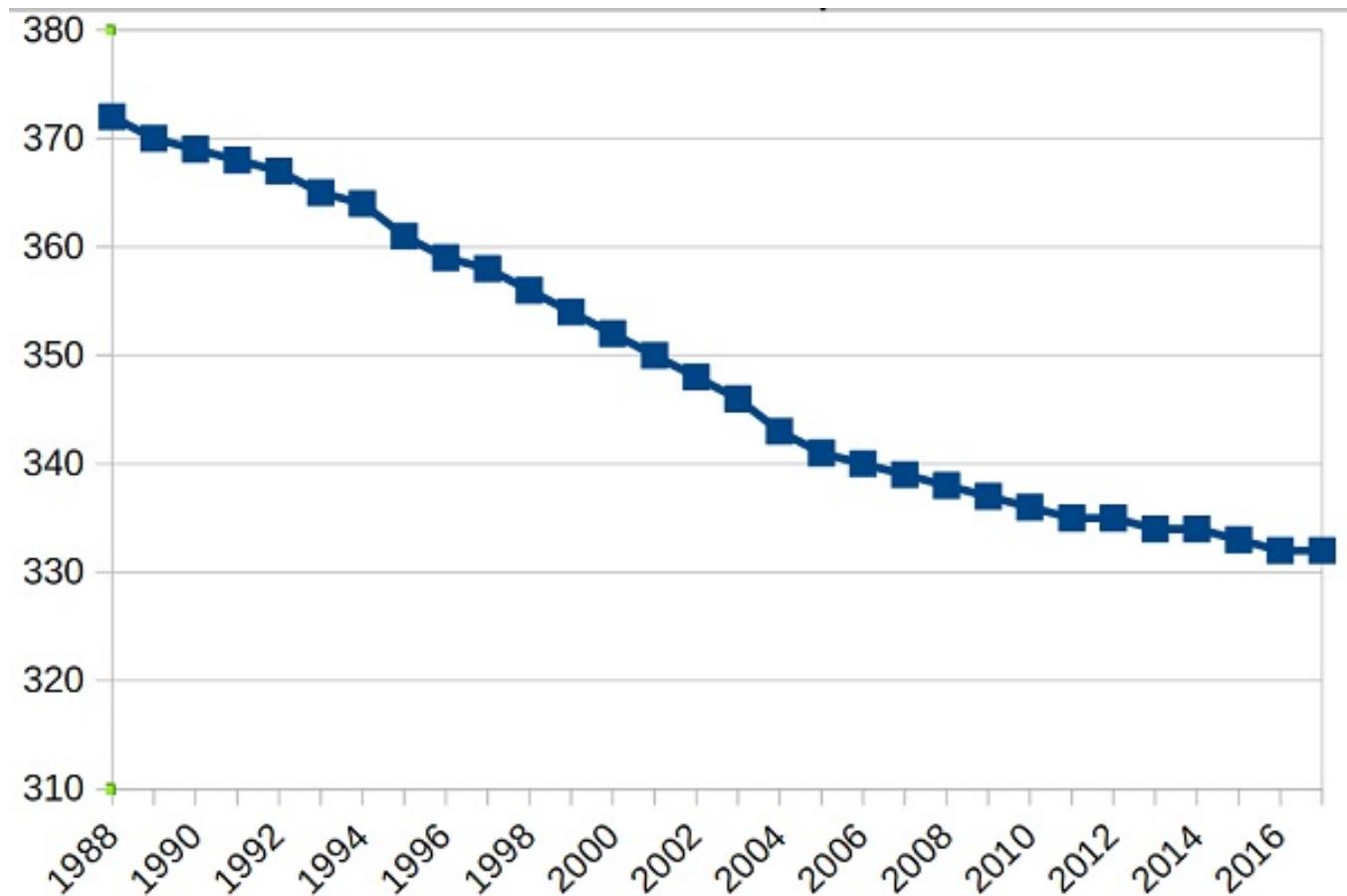
Or maybe it was the declining-rainforest-coverage-over-time-maps. I feel like about one in every three posters I saw as a child looked something like this:



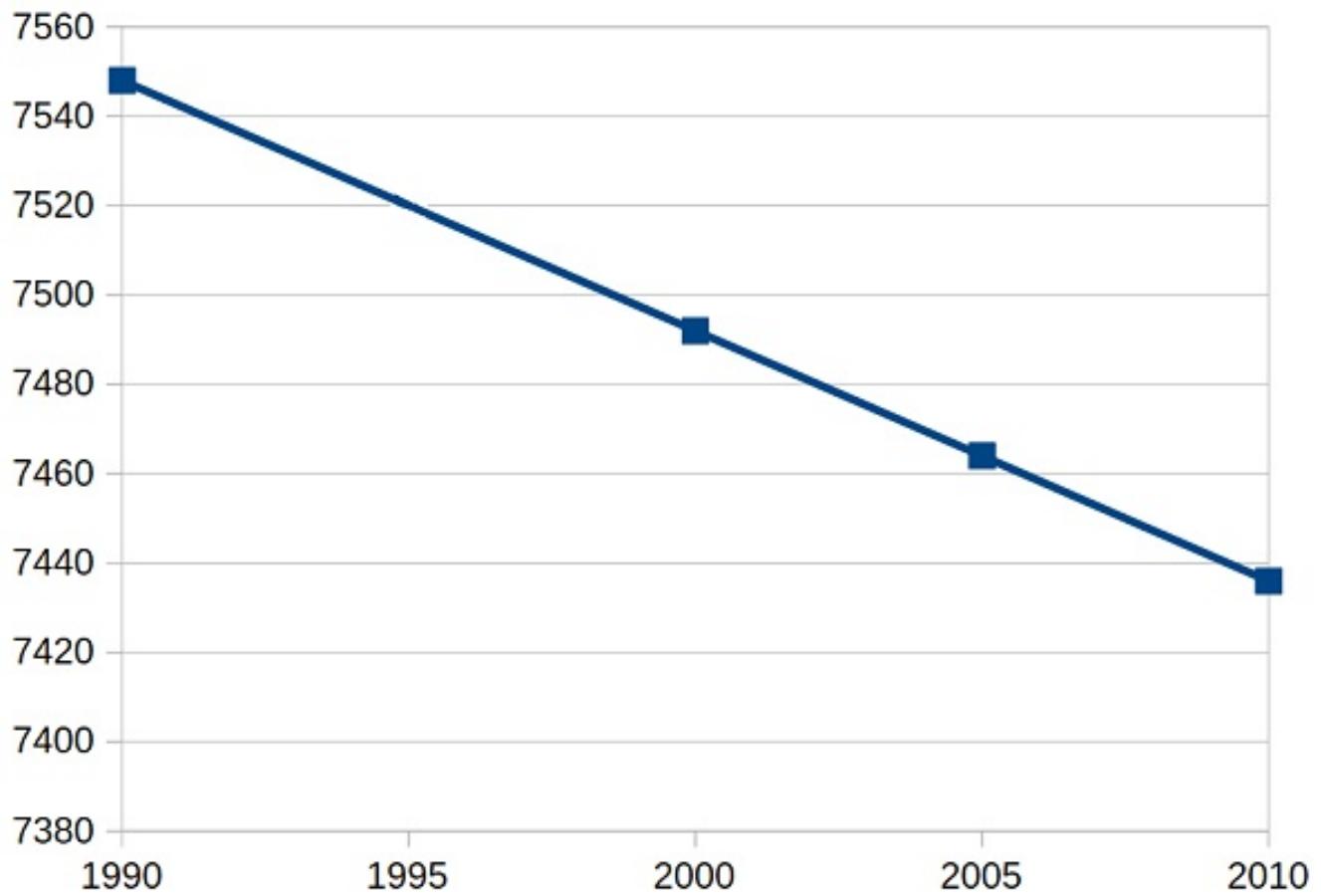
This is a fake example. Please stop asking me where I am getting the data from.

I thought surely nothing could be easier than digging up a few of them and seeing whether their 2020 predictions were right. But I can't find them anywhere. According to the Internet, there is no such thing as 90s-era maps showing declining rainforest coverage over time. Can anyone else locate these?

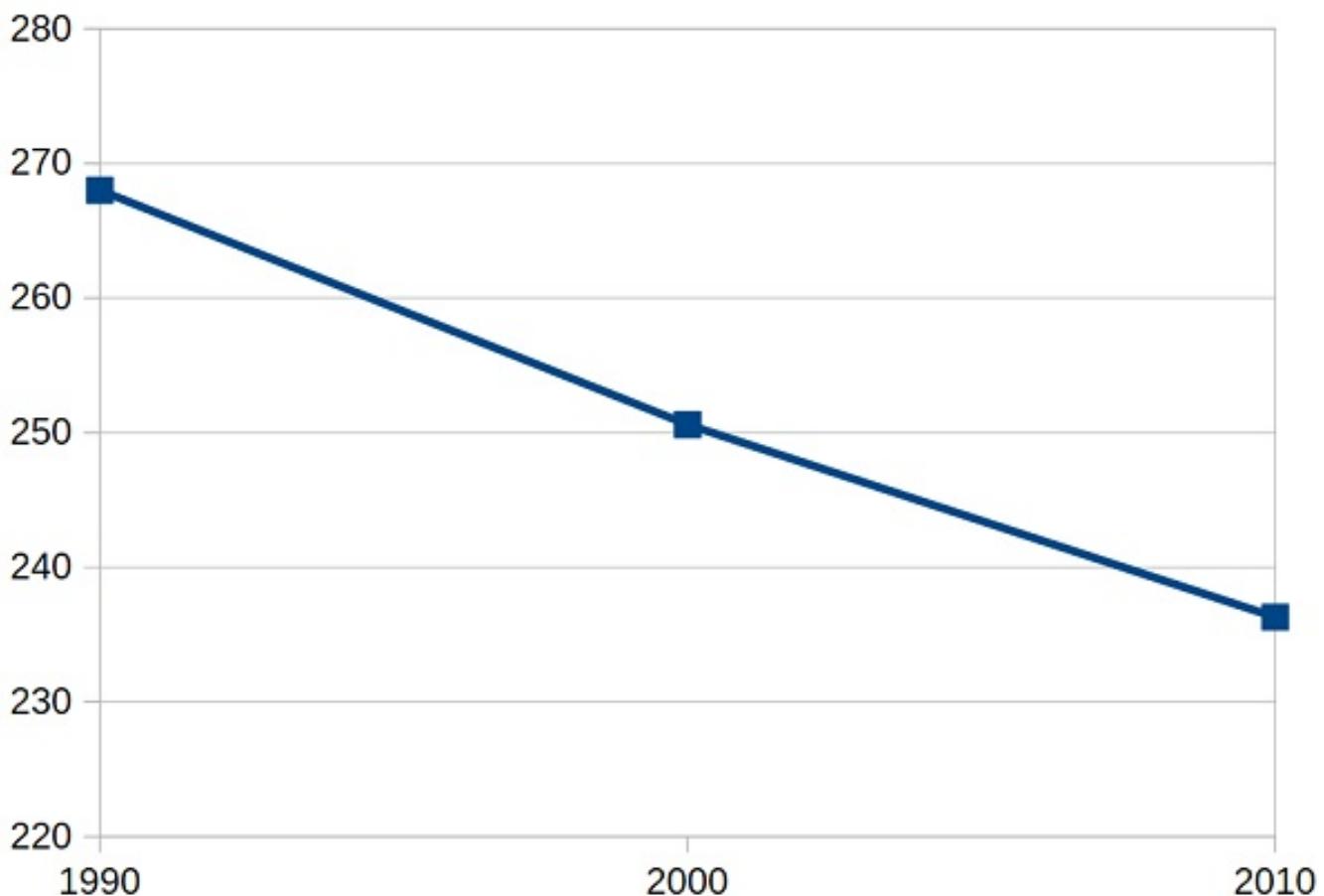
Anyway:



Here's a graph of the size of the Amazon over time ([source](#), note that the y-axis is not at zero). At 90s levels of deforestation, the Amazon would have disappeared in about 200 years. At current levels, it will disappear in about 400 years.



Here's the Congo (somewhat dubious [source](#), same caveat). At the rates shown here it will be gone in 250 years – but it seems to have [slowed](#) after the period on the graph.



And here's Southeast Asia ([source](#), same caveat). At this rate, the southeast Asian forest will be gone in 150 years, though some new papers are suggesting [we may be underestimating the deforestation rate](#).

Overall it looks like deforestation may have decreased modestly in the Amazon (and possibly the Congo) since the 1990s. It has not decreased significantly in Southeast Asia, and whatever decreases have happened are not relevant to the scale of the problem.

The only good news is that all those "rainforests will be gone by 2050" posters were just wrong; there is more rainforest than that. But not *that much* more.

Verdict: *The problem still exists, and we are just ignoring it now.*

4. Endangered Species

So just find how many species go extinct each year, and whether it's a lot or a little, and then we'll know what's going on with this, right? Ha ha, as if.

On the one hand, the UN Environment Programme says that "150-200 species of plant, insect, bird and mammal become extinct every 24 hours."

On the other, nobody can name more than a single-digit number of species that go extinct in any given year. The [2017 list](#) includes five: a bat, a cat, a flatworm, a lizard, and a snail. This matches longer-term surveys: [Ceballos et al \(2018\)](#)

find that about 477 vertebrate species have gone extinct since 1900 – again, about five per year. And a recent survey found only [four to eight bird species](#) had disappeared since the turn of the century.

I have no idea where the 150-200 number per day comes from, and neither does anyone else. The closest I can find to a justification is [this WWF page](#), which reminds us that if there are 100 million animals species, and “the extinction rate is just 0.01% per year”, then *at least* 10,000 species go extinct every year (=200-300/day) – but all of these numbers are completely made up.

One could try to justify these estimates with something like “assume only one in a thousand species has been discovered and is monitored well enough to detect its extinction, so if we detect five extinctions per year then five thousand must be happening” – but I’ve never heard anyone actually say this. Also, with apologies to all the undiscovered species, if they’re so tiny and uncommon as to never get discovered, it doesn’t seem like their extinction is going to change very much.

Five known species going extinct per year may sound like a lot if you’re thinking it’s something like “rhinos, pandas, whales, spotted owls, and leopards”. But realistically there are 385 species of shrews. We could spend our entire yearly extinction budget on shrews for the next sixty years and still have more than enough kinds of shrews left to satisfy basically anybody.

I’m trying to think what the best counterargument to this would be – the best case that we really do need to consider species extinction a dire concern.

Maybe this is too vertebrate-centric, and there are lots of insects and plants and such going extinct all the time? But this [List Of Recently Extinct Insects](#) suggests that of about 6000 known insect species, only 50-100 have gone extinct in the past century. And one of those was [this giant earwig](#) which I really think the world is better without.

Or maybe we can’t directly predict the future from the past. Imagine 1000 square miles of rainforest with a homogenous distribution of species. Clear-cut 50% of the rainforest, and no extinctions. Clear-cut 90% of the rainforest, still no extinctions. Clear-cut 99%, maybe a few extinctions if you’re unlucky. Then clear-cut the last remaining 1% and everything dies. It seems like something like that might be happening – see for example [this report](#) that global animal populations have declined 58% over the past forty years.

But any concept of endangered species that focuses on “many well-known species will be gone soon” doesn’t seem consistent with the evidence.

Verdict: Partly alarmism, partly still going on.

5. More And More Trash Piling Up Until The Whole World Is Just A Giant Mountain Of Trash

Wait, what? Was this really a concern? Did I really spend my primary school years being told that if I didn’t vigilantly recycle everything, one day I would be submerged beneath a sea of trash, breathing by means of a trash snorkel? Am I hallucinating all of this?

As usual, it turns out to be the Mafia’s fault. In the 1980s, mob boss Salvatore Avellino took over New York City’s landfill industry, and in a shocking development which nobody could have predicted, was corrupt. New York City soon ran out of landfill space. Somehow all of its excess trash ended up on a barge called the MOBRO-4000, because the Eighties, and this barge apparently sailed up and down the east coast of North America searching for a place to deposit

its trash. In its many exciting adventures it reached the coast of Belize, got involved in a confrontation with the Mexican Navy, and finally went back to New York, where at some point landfill space was found and the crisis was over.

But a giant boat full of trash made a really memorable image, and it got nationwide news coverage, and environmentalists took advantage of this to tell everyone there was no more landfill space anywhere in the world and we all had to recycle right now. According to [Wikipedia](#):

At the time, the *Mobro 4000* incident was widely cited by environmentalists and the media as emblematic of the solid-waste disposal crisis in the United States due to a shortage of landfill space: almost 3,000 municipal landfills had closed between 1982 and 1987. It triggered much national public discussion about waste disposal, and may have been a factor in increased recycling rates in the late 1980s and after. It was this that caused it to be included in an episode of *Penn & Teller: Bullshit!* (season 2, episode 5) in which they debunk many recycling myths.

I'm even absolutely right in remembering primary school lessons centered around garbage covering the Earth and killing everybody. [Here's](#) a *New York Times* article from 1996 – ie after the crisis had a little bit of time to fade – lightly mocking the new curricula that followed in its wake:

After the litter hunt in Ms. Aponte's science classroom, it was time for a guest lecturer on garbage. A fifth-grade class was brought in to hear Joanne Dittersdorf, the director of environmental education for the Environmental Action Coalition, a nonprofit group based in New York. Her slide show began with a 19th-century photograph of a street in New York strewn with garbage.

"Why can't we keep throwing out garbage that way?" Dittersdorf asked.

"It'll keep piling up and we won't have any place to put it."

"The earth would be called the Trash Can."

"The garbage will soon, like, take over the whole world and, like, kill everybody."

Dittersdorf asked the children to examine their lives. "Does anyone here ever have takeout food?" A few students confessed, and Dittersdorf gently scolded them. "A lot of garbage there."

She showed a slide illustrating New Yorkers' total annual production of garbage: a pile big enough to fill 15 city blocks to a height of 20 stories. "There are a lot of landfills in New York City," Dittersdorf said, "but we've run out of space."

From the same beginning-of-the-backlash period we also get this 1995 Foundation for Economic Education piece, [Are We Burying Ourselves In Garbage?](#)

A popular idea in public discourse today is that the United States produces an overwhelming amount of trash—so much that our landfills will not be able to handle the quantity. The most eloquent symbol of this viewpoint was the “garbage barge,” which in the late 1980s left Long Island and could not find a port or country willing to accept its 3.168 tons of refuse. [But] the actual data (such as they are) on the amount of municipal solid waste produced present us with more questions than answers.

This article also deserves note for hitting on a brilliant solution:

The crisis mentality has distorted judgment of waste disposal. The notion that modern America is especially wasteful is demonstrably wrong, both in terms of the last decades as well as the last 100 years. The idea that our landfills are literally “running out” is even less credible. If in the next century major portions of the United States really need to export their refuse to other states, a “gold mine” for refuse burial does exist: South Dakota. This state is geologically, economically, and politically almost ideal for massive municipal solid waste management.

None of this is a joke. This is how your parents did Discourse, people.

But it turns out capitalism works: if there’s a shortage of landfills, that incentivizes people to create new landfills. Also, the world is very large and it is hard to cover a significant portion of it in trash. There was a brief blip as cities figured out how to pay for more waste disposal, and then nobody ever worried about the problem again. Recycling remained inefficient and of dubious benefit, and never really caught on.

There is still an international problem as Third World countries struggle with infrastructure issues around trash disposal. You still see occasional articles like Huffington Post’s [People Are Living In Landfills As The World Drowns In Its Own Trash](#), from earlier this year. But I think in general nobody in the First World still considers this a major problem.

Well, *almost* nobody:

The earth, our home, is beginning to look more and more like an immense pile of filth.

— [Pope Francis \(@Pontifex\)](#), June 18, 2015

Verdict: Alarmist. So, so, alarmist.

6. Peak Resource

Is the earth’s ballooning human population using up resources at an unsustainable rate?

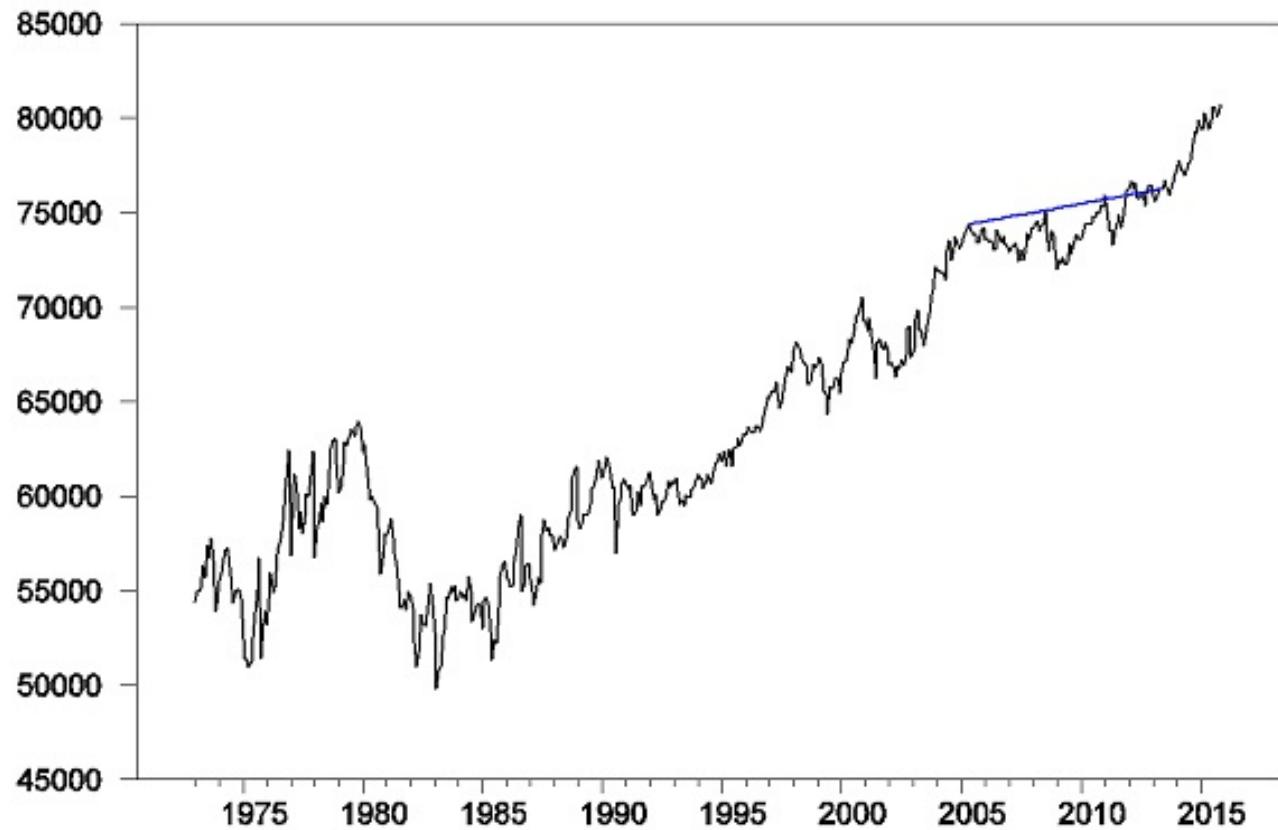
Technically the answer must be “yes”, since by definition nonrenewable resources have to run out at some point. But when? Long after we have escaped to space and gotten access to shiny new resources? Or soon enough that we have to worry about it?

A big part of 90s environmentalism involved worrying that it was the latter. A particular concern was “peak oil”, the point at which we had exhausted so much of the world’s oil that production rates declined every year thereafter and oil started becoming gradually rarer and more expensive. Wikipedia has a helpful table of people’s peak oil predictions. I’ve highlighted the ones that have already passed in red.

Pub.	Made by	Peak year/range	Pub.	Made by	Peak year/range
1972	Esso	About 2000	1999	Parker	2040
1972	United Nations	By 2000	2000	A. A. Bartlett	2004 or 2019
1974	Hubbert	1991–2000	2000	Duncan	2006
1976	UK Dep. of Energy	About 2000	2000	EIA	2021–2067; 2037 most likely
1977	Hubbert	1996	2000	EIA (WEO)	Beyond 2020
1977	Ehrlich, et al.	2000	2001	Deffeyes	2003–2008
1979	Shell	Plateau by 2004	2001	Goodstein	2007
1981	World Bank	Plateau around 2000	2002	Smith	2010–2016
1985	J. Bookout	2020	2002	Campbell	2010
1989	Campbell	1989	2002	Cavallo	2025–2028
1994	L. F. Ivanhoe	OPEC plateau 2000–2050	2003	Greene, et al.	2020–2050
1995	Petroconsultants	2005	2003	Laherrère	2010–2020
1997	Ivanhoe	2010	2003	Lynch	No visible peak
1997	J. D. Edwards	2020	2003	Shell	After 2025
1998	IEA	2014	2003	Simmons	2007–2009
1998	Campbell & Laherrère	2004	2004	Bakhitari	2006–2007
1999	Campbell	2010	2004	CERA	After 2020
1999	Peter Odell	2060	2004	PFC Energy	2015–2020

A selection of estimates of the year of peak world oil production, compiled by the United States Energy Information Administration

Almost everyone working before 2000 thought we would have reached peak oil by now. But here's world oil production over time:



And the price of oil:



What happened? People discovered fracking and other paradigm-shifting techniques to extract oil from shale, which opened up vast new previously-inaccessible oil fields. The peak oil predictors might call this unfair – they calculated correctly given the technology they knew about – but the *whole argument* of the people who say we don't have to worry about peak resource (sometimes called "cornucopians") is that technology will advance fast enough to satisfy our resource needs. In this case they were right.

What about non-oil resources?

In 1980, leading environmental scientist and peak-resource proponent Paul Ehrlich made a bet with cornucopian economist Julian Simon about how resource prices would change over the next decade. The [Simon-Ehrlich Wager](#) has become a famous example of futurology done right – two people with different theories implying different predictions coming together, agreeing on exactly what each of their theories implied, and then publicly committing to put them to the test. According to Reb Wiki:

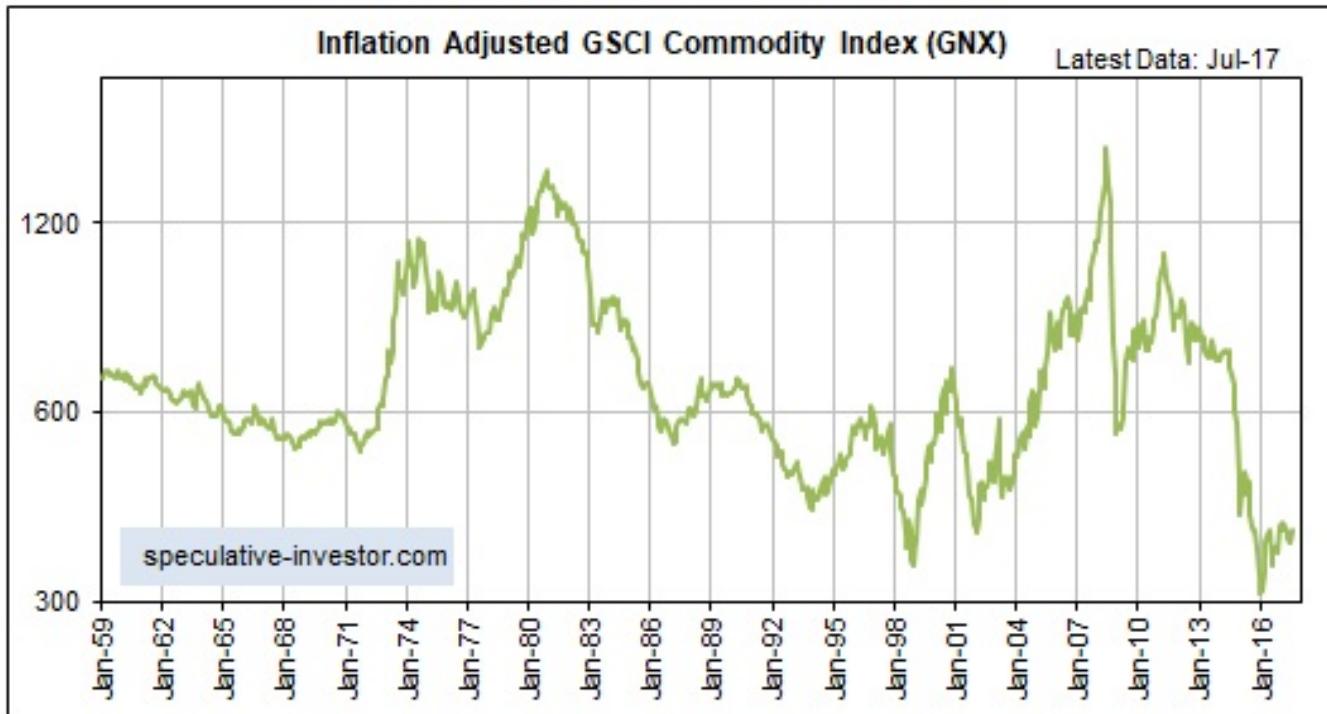
Simon challenged Ehrlich to choose any raw material he wanted and a date more than a year away, and he would wager on the inflation-adjusted prices decreasing as opposed to increasing. Ehrlich chose copper, chromium, nickel, tin, and tungsten. The bet was formalized on September 29, 1980, with September 29, 1990 as the payoff date. Ehrlich lost the bet, as all five commodities that were bet on declined in price from 1980 through 1990, the wager period.

Looks pretty good for Simon and the cornucopians. But the article continues:

Ehrlich could have won if the bet had been for a different ten-year period. Ehrlich wrote that the five metals in question had increased in price between the years 1950 to 1975. Asset manager Jeremy Grantham wrote that if the Simon–Ehrlich wager had been for a longer period (from 1980 to 2011), then Simon would have lost on four of the five metals. He also noted that if the wager had been expanded to "all of the most important commodities," instead of just five metals, over that longer period of 1980 to 2011, then Simon would have lost "by a lot." Economist Mark J. Perry noted that for an even longer period of time, from 1934 to 2013, the inflation-adjusted price of the Dow Jones-AIG Commodity Index showed "an overall significant downward trend" and concluded that Simon was "more right than lucky". Economist Tim Worstall wrote that "The end result of all of this is that yes, it is true that Ehrlich could have, would have, won the bet depending upon the starting date. ... But the long term trend for metals at least is downwards."

How about today? [An econblogger](#) is still keeping track of the Ehrlich-Simon wager, and finds that as of August 2017, Simon (who is now dead) is still winning; a basket of the five metals involved still costs less than it did in 1980.

Can we zoom out even further? There are a bunch of commodity indices that do for commodities what the Dow Jones does for stocks. I chose the Standard & Poor Goldman-Sachs Commodity Index kind of randomly because they were a familiar name and it was easy to find [which goods](#) they included. I'm not quite sure I'm doing this right, but this seems to be the most relevant graph:



The price of commodities in general is still lower than in 1980 (also, with this graph it becomes clear Ehrlich was really unlucky in which year he started his wager).

I have never heard anyone claim that this represents an environmentalist victory: I don't think there was any large-scale attempt to conserve or recycle chromium/tungsten/whatever that led to its current abundance. I think this was just a victory for resource extraction technology.

There are still theoretical reasons to think we have to run out of stuff eventually. But in terms of how the past 25 years have treated 90s-era concerns about resource depletion, it's hard to answer anything other than "savagely".

Verdict: Alarmist.

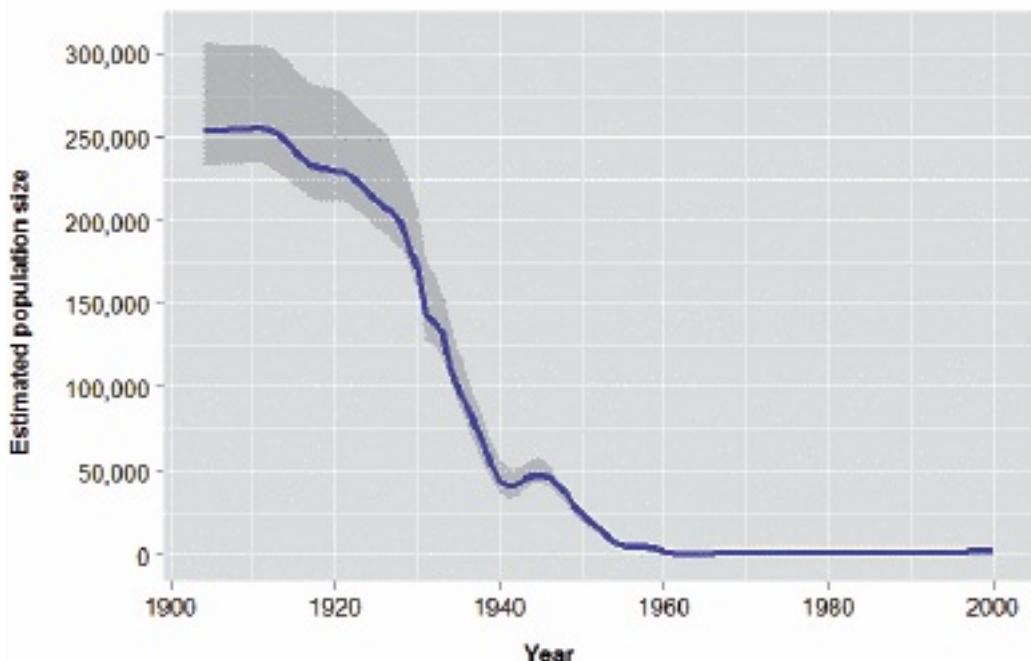
7. Saving The Whales

I remember frequently being told we had to do this. Apparently it paid off, since a global moratorium on whaling was signed in 1982.

The ban is not perfect. Indigenous peoples are allowed to hunt whales in traditional ways. Japan pretends their whaling is for "scientific purposes" and has so far gotten away with it. Norway and Iceland never signed the moratorium and continue to whale.

But overall, things are going pretty well. There aren't a lot of graphs, but the [International Whaling Commission](#) (which despite its name is against whaling) says blue whale populations are increasing at about 8%/year, humpback whales around 10%, and fin whales around 5%. Those sound pretty good, but they have to be taken in context:

Decline and slow recovery of Antarctic blue whales



Okay, fine. There's one graph. But it's really depressing. See that tiny micro-bump at the end? That represents progress.

Verdict: Environmental movement successfully solved this problem.

8. Concluding Thoughts

This was not a very conclusive exercise. When I add these up – as if that were at all an acceptable thing to do! – I get 2½ that were solved, 2% that were alarmism, and 1% that continue. So there is not much to be said about them as a group. Some were solved through heroic effort. Some turned out to be completely made up. Some of them are still out there but have stopped capturing the public's attention.

Victories I can understand. It's the latter two categories that confuse me.

How did the non-problems fade away? There was no moment when some brave iconoclast posted ninety-five theses to the door of the local recycling center and said "No! There is not a landfill crisis!" I mean, John Tierney wrote things along those lines, and did a great job of it. But he's not a household name and there was never a time when everyone said "Oh, John Tierney is right, let's stop worrying about this." The people who stopped worrying about this never heard of John Tierney. At some point people just went from being very worried about the landfill crisis to shaking their heads and saying "The world getting full of trash? Sounds pretty stupid."

And the story with peak resources seems entirely different. You will still occasionally see people saying "The Earth can't support our greed, soon we will run out of everything", and reasonable people will nod along with this and admit it is very wise. But you hear it like once a year now, as opposed to it being a constant refrain. This idea was never intellectually defeated at all, at least not on the popular level. It just faded away.

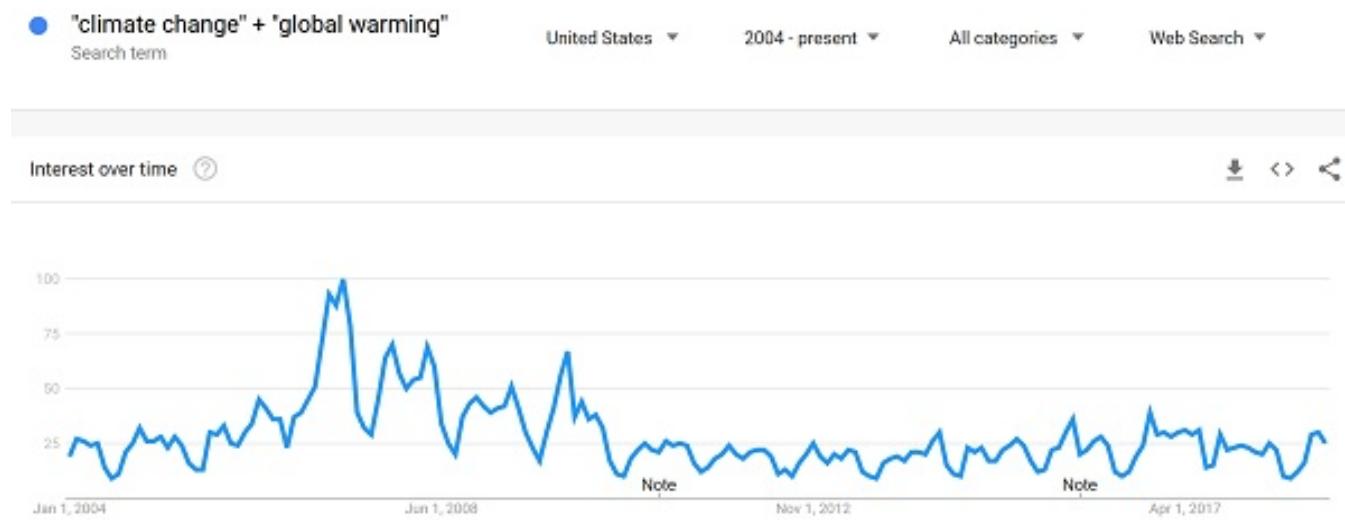
Was there some rarified level of intellectual debate where these ideas lost out? And then, denied their support from the commanding heights of the ivory tower, did journalists stop writing about them, schoolteachers stop teaching them, and then eventually the public – who have no will of their own and have to be told what matters – wander off and do something else?

Or was the change bottom-up? Did the public, after the millionth editorial on the trash crisis, say “Okay, whatever”, such that journalists realized this was no longer a good way to sell newspaper subscriptions? Is there a natural news mega-cycle of a decade or so, after which the public gets tired of hearing about a certain story, the intellectuals get tired of talking about it, every possible angle has been explored, and people move on, whether or not it was solved? Does this explain why the rainforests, a real problem that is still going on, similarly lost public attention?

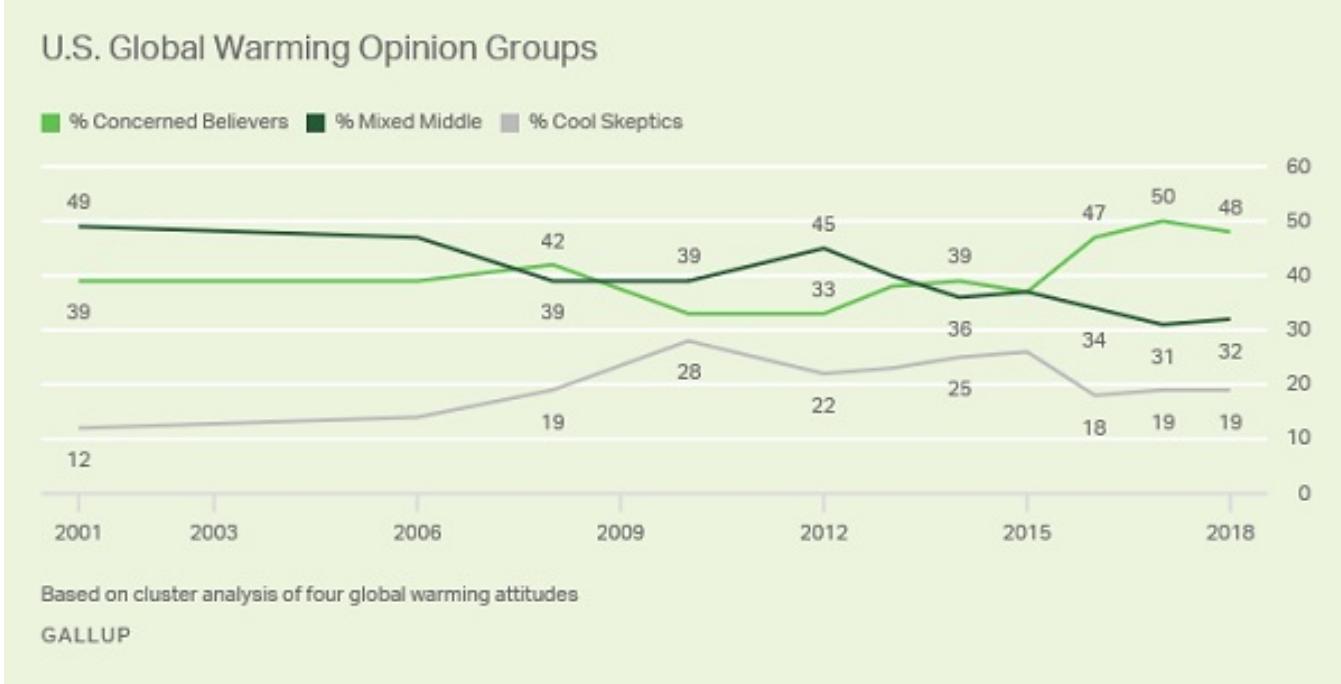
Or maybe climate change took over everything, became so important that everything else faded into the background. This is certainly how it feels to me. Whenever I hear about the rainforests nowadays, it's as a footnote to some global warming story where they add that we should save the rainforest as a carbon sink. Whenever I hear about landfills or recycling today, it's in the context of trash giving off greenhouse gases. It feels almost like some primitive barter system has been converted to a modern economy, with tons of CO₂ emission as the universal interchangeable currency that can be used to put a number value on all environmental issues. Can't figure out a way to convert whales into a carbon sink? Guess they'll have to go.

(I wrote that, then remembered I lived in 21st century America, did a Google search, and sure enough there are dozens of articles arguing that saving whales is an efficient way to neutralize greenhouse gases)

But as attractive as this picture is, it's hard to find the supporting data. There's just not hard evidence that we care more about global warming than we did fifteen or twenty years ago:



Here's the Google Trends. There was a lot of interest in 2006, which I think gets attributed to Al Gore's *An Inconvenient Truth* in that year, but not a lot of signs of increase today.



Here's Gallup. It at least shows a spike starting in 2016 – but given its timing and the lack of obvious 2016 global-warming related events, I think it's probably just another [Trump backlash effect](#).

If global warming is eating all the other environmental issues, it doesn't seem to be extracting that much nutrition from their corpses. And the ozone hole – probably the most global-warming-like issue of the last generation – managed to gather popular support at the same time that people were worried about a host of other things. I don't know. Maybe given the public's tendency to get bored of an issue after a decade or so, global warming has to cannibalize the rest of environmentalism just to survive at all. Depressing if true.

Or maybe it's a *zeitgeist* thing. For some reason, it's hard to imagine 2018 being the Year Of Rainforest Concern. There's something very 90s Optimism about worrying about the rainforests, something where even the warnings of doom have a cheerful ring to them. I remember a Rainforest Charity Box at my local mall as a kid, promising that if you donated \$10, you would save a brightly colored parrot, and if you donated \$50, you might save a jaguar. Who thinks that way these days? Now if you donate some amount to stopping global warming, you will have won yourself [a lecture](#) from a bunch of people telling you that still doesn't mean you have the right to feel good about yourself, and the world is going to fry regardless. Have we just passed the point where anybody can care about crisp mountain streams or frolicking snow leopards any more?

The most important thing I take away from the exercise is a sort of postmodern insight into the way environmental issues are constructed. This is *definitely not* me saying they are all made up; many of them are very real. But the mapping from real crisis to social panic is tenuous, contingent, and historical. Sometimes random things that shouldn't matter get magnified into the *issue du jour*; other times giant world-threatening crises manage to slip everyone's attention.

Imagine that twenty years from now, nobody cares or talks about global warming. It hasn't been debunked. It's still happening. People just stopped considering it interesting. Every so often some webzine or VR-holozine or whatever will publish a "Whatever Happened To Global Warming" story, and you'll hear that global temperatures are up X degrees centigrade since 2000 and that explains Y percent of recent devastating hurricanes. Then everyone will go back to worrying about Robo-Trump or Mecha-Putin or whatever.

If this sounds absurd, I think it's no weirder than what's happened to 90s environmentalism and the issues it cared about.

The Atomic Bomb Considered As Hungarian High School Science Fair Project

Posted on May 26, 2017 by Scott Alexander



I

A group of Manhattan Project physicists [created](#) a tongue-in-cheek mythology where superintelligent Martian scouts landed in Budapest in the late 19th century and stayed for about a generation, after which they decided the planet was unsuitable for their needs and disappeared. The only clue to their existence were the children they had with local women.

The joke was that this explained why the Manhattan Project was led by a group of Hungarian supergeniuses, all born in Budapest between 1890 and 1920. These included Manhattan Project founder [Leo Szilard](#), H-bomb creator [Edward Teller](#), Nobel-Prize-winning quantum physicist [Eugene Wigner](#), and legendary polymath [John von Neumann](#), namesake of [the List Of Things Named After John Von Neumann](#).

The coincidences actually pile up beyond this. Von Neumann, Wigner, and possibly Teller all went to the same central Budapest high school at about the same time, leading a friend to joke about the atomic bomb being *basically* a Hungarian high school science fair project.

But maybe we shouldn't be joking about this so much. Suppose we learned that Beethoven, Mozart, and Bach all had the same childhood piano tutor. It sounds less like "ha ha, what a funny coincidence" and more like "wait, who was this guy, and how quickly can we make everyone else start doing what he did?"

In this case, the guy was [Laszlo Ratz](#), legendary Budapest high school math teacher. I didn't even know people *told* legends about high school math teachers, but apparently they do, and this guy features in a lot of them. There is apparently a Laszlo Ratz Memorial Congress for high school math teachers each year, and a Laszlo Ratz medal for services to the profession. There are plaques and statues to this guy. It's pretty impressive.

A while ago I looked into the literature on teachers [and concluded](#) that they didn't have much effect overall. Similarly, Freddie deBoer writes that most claims that certain schools or programs have transformative effects on their students [are the result of selection bias](#).

On the other hand, we have a Hungarian academy producing like half the brainpower behind 20th century physics, and Nobel laureates who literally [keep a picture](#) of their high school math teacher on the wall of their office to inspire them. Perhaps even if teachers don't explain much of the *existing* variability, there are heights of teacherdom so rare that they don't show up in the statistics, but still exist to be aspired to?

II

I've heard this argument a few times, and I think it's wrong.

Yes, two of Ratz's students went on to become supergeniuses. But Edward Teller, another supergenius, went to the same high school but (as far as I know) was never taught by Ratz himself. That suggests that the school was good at producing supergeniuses regardless of Ratz's personal qualities. A further point in support of this: [John Harsanyi](#) also went to the school, also wasn't directly taught by Ratz, and also went on to win a Nobel Prize and invent various important fields of mathematics. So this school – the [Fasori Gymnasium](#) – seems to have been about equally excellent for both its Ratz-taught and its non-Ratz-taught pupils.

Yet the Fasori Gymnasium *might not have even been the best high school in its neighborhood*. It competed with the Minta Gymnasium half a mile down the street, whose alumni include Manhattan Project physicists [Nicholas Kurti](#) and [Theodore von Karman](#) (von Karman went on to found the Jet Propulsion Laboratory), brilliant chemist-philosopher [Michael Polanyi](#), economists [Thomas Balogh](#) and [Nicholas Kaldor](#) (of Kaldor-Hicks efficiency fame), and [Peter Lax](#), who [once said](#) "You don't have to be Hungarian to be a mathematician – but it helps". There are also some contradictory sources suggesting Teller attended this school and not Fasori; for all I know he might have attended both. Once again, most of these people were born in the 1890-1910 period when the Martian scouts were supposedly in Budapest.

Worse, I'm not even sure that the best high school in early 20th-century Hungary was *either* of the two mentioned above. The Berzsenyi Gymnasium, a two mile walk down Gyorgy Street from the others, boasts alumni including multizillionaire [George Soros](#), Intel founder [Andrew Grove](#), BASIC inventor [John Kemeny](#), leading cancer biologist [George Klein](#), great mathematician [George Polya](#), and Nobel Prize winning physicist [Dennis Gabor](#).

Given that the Fasori Gymnasium wasn't obviously better than either of these others, is it possible that the excellence was at a higher level – neither excellent teachers nor excellent principals, but some kind of generally excellent Hungarian culture of education?

This is definitely what the Hungarians want us to think. According to [Cultures of Creativity](#):

What's so special about Budapest's schools? A certain elitism and a spirit of competition partly explains the successes of their students. For example, annual competitions in mathematics and physics have been held since 1894. The instruction the students receive as well as these contests are an expression of a special pedagogy and a striving to encourage creativity. Mor Karman, founder of the Minta school, believed that everything should be taught by showing its relation to everyday life. Instead of learning rules by heart from books, students tried to formulate the rules themselves.

This paper on ["The Hungarian Phenomenon"](#) makes similar claims, but adds a few more details:

The Eotvos Contests were a powerful mean for the stimulation of mathematics on a large scale and were used to motivate mathematical culture in the society. It also provided a channel to search for talented youths. The contests, which have been open to Hungarian high school students in their last year since 1894, played a remarkable role in the development of mathematics.

Okay. But I want to challenge this. During this era, formal education in Hungary began at age 10. By age ten, John von Neumann, greatest of the Hungarian supergeniuses, already spoke English, French, German, Italian, and Ancient Greek, knew integral and differential calculus, and could multiply and divide 8-digit numbers in his head. Wikipedia notes that on his first meeting with his math teacher, the math teacher "was so astounded with the boy's mathematical talent that he was brought to tears". This doesn't sound like a guy whose potential was kindled by formal education. This sounds like a guy who would have become one of history's great mathematicians even if his teachers had slept through his entire high school career.

Likewise, the book above notes that Dennis Gabor, the Hungarian inventor of holography, "developed his passion for

physics during his youth, but did so for the most part on his own". [His biography](#) notes that "During his childhood in Budapest, Gabor and his brother would often duplicate the experiments they read about in scientific journals in their home laboratory."

Likewise, consider [Paul Erdos](#), a brilliant mathematician born in Budapest around this time. As per his Wikipedia page, "Left to his own devices, he taught himself to read through mathematics texts that his parents left around their home. By the age of four, given a person's age, he could calculate, in his head, how many seconds they had lived."

I have no knock-down proof that Hungary's clearly excellent education system didn't contribute to this phenomenon. A lot of child prodigies burn out, and maybe Hungary was unusually good at making sure that didn't happen. But it sure seems like they had a lot of child prodigies to work with.

So what's going on? Should we just accept the Manhattan Project consensus that there was a superintelligent Martian scout force in early 20th-century Budapest?

III

Here's something interesting: every single person I mentioned above is of Jewish descent. *Every single one*. This isn't some clever setup where I only selected Jewish-Hungarians in order to spring this on you later. I selected all the interesting Hungarians I could find, then went back and checked, and every one of them was Jewish.

This puts the excellence of the Hungarian education system in a different light. Hungarian schools totally failed to work their magic on Gentiles. You can talk all you want about "elitism and a spirit of competition" and "striving to encourage creativity", yet for some reason this worked on exactly one of Hungary's many ethnic groups.

This reduces the difficult question of Hungarian intellectual achievement to the easier question of Jewish intellectual achievement.

I say "easier question" because I find the solution by Cochran, Hardy, and Harpending really compelling. Their paper is called [A Natural History Of Ashkenazi Intelligence](#) ("Ashkenazi" means Eastern European Jew) and they start by expressing the extent of the issue:

Ashkenazi Jews have the highest average IQ of any ethnic group for which there are reliable data. They score 0.75 to 1.0 standard deviations above the general European average, corresponding to an IQ 112 – 115. This fact has social significance because IQ (as measured by IQ tests) is the best predictor we have of success in academic subjects and most jobs. Ashkenazi Jews are just as successful as their tested IQ would predict, and they are hugely overrepresented in occupations and fields with the highest cognitive demands. During the 20th century, they made up about 3% of the US population but won 27% of the US Nobel science prizes and 25% of the Turing Awards [in computer science]. They account for more than half of world chess champions.

This doesn't seem to be due to any advantage in material privilege; Ashkenazi Jews frequently did well even in countries where they were persecuted. Nor is it obviously linked to Jewish culture; Jews from other regions of the world show no such advantage. So what's going on?

Doctors have long noted that Ashkenazi Jews are uniquely susceptible to various genetic diseases. For example, they're about a hundred times more likely to have [Gaucher's Disease](#), a hundred times more likely to get [Tay-Sachs Disease](#), ten times more likely to have [torsion dystonia](#), et cetera. Genetic diseases are so common in this population

that the are [official recommendation](#) is that *all* Ashkenazi Jewish couples get screened for genetic disease before marriage. I'm Ashkenazi Jewish, I got screened, and I turn out to be a carrier for [Riley-Day syndrome](#) – three hundred times as common in Ashkenazi Jews as in anyone else.

Evolution usually gets rid of genetic diseases pretty quickly. If they stick around, it's because they're doing something to earn their keep. One common pattern is "[heterozygote advantage](#)" – two copies of the gene cause a disease, but one copy does something good. For example, people with two copies of the sickle cell gene get sickle cell anaemia, but people with one copy get some protection against malaria. In Africa, where malaria is relatively common, the tradeoff is worth it – so people of African descent have high rates of the sickle cell gene and correspondingly high rates of sickle cell anaemia. In other places, where malaria is relatively uncommon, the tradeoff isn't worth it and evolution eliminates the sickle cell gene. That's why sickle cell is about a hundred times more common in US blacks than US whites.

The moral of the story is: populations can have genetic diseases if they also provide a useful advantage to carriers. And if those genetic diseases are limited to a single group, we expect them to provide a useful advantage for that group, but not others. Might the Jewish genetic diseases provide some advantage? And why would that advantage be limited to Jews?

Most of the Jewish genetic diseases cluster into two biological systems – the sphingolipid system and the DNA repair system. This is suspicious. It suggests that they're not just random. They're doing something specific. Both of these systems are related to neural growth and neural branching. Might they be doing something to the brain?

Gaucher's disease, one of the Ashkenazi genetic diseases, appears to increase IQ. CHH obtained a list of all of the Gaucher's patients in Israel. They were about 15 times more likely than the Israeli average to be in high-IQ occupations like scientist or engineer; CHH calculate the probability that this is a coincidence to be $4 * 10^{-19}$.

Torsion dystonia, another Ashkenazi genetic disease, shows a similar pattern. CHH find ten reports in the literature where doctors comment on unusual levels of intelligence in their torsion dystonia patients. [Eldridge, Harlan, Cooper, and Riklan](#) tested 14 torsion dystonia patients and found an average IQ of 121; another similar study found an average of 117. Torsion dystonia is pretty horrendous, but sufferers will at least get the consolation prize of being really, really smart.

Moving from medicine to history, we find that Ashkenazi Jews were persecuted for the better part of a millennium, and the particular form of this persecution was locking them out of various jobs until the main career opportunities open to them were things like banker, merchant, and doctor. CHH write:

For 800 to 900 years, from roughly 800 AD to 1650 or 1700 AD, the great majority of the Ashkenazi Jews had managerial and financial jobs, jobs of high complexity, and were neither farmers nor craftsmen. In this they differed from all other settled peoples of which we have knowledge.

They continue:

Jews who were particularly good at these jobs enjoyed increased reproductive success. Weinryb (1972, see also Hundert 1992) comments: "More children survived to adulthood in affluent families than in less affluent ones. A number of genealogies of business leaders, prominent rabbis, community leaders, and the like – generally belonging to the more affluent classes – show that such people often had four, six, sometimes even eight or nine children who reached adulthood. On the other hands, there are some indications that poorer families tended to be small ones... as an example, in a census of the town of Brody in 1764 homeowner households had 1.2 children per adult member while tenant households had 0.6."

Now we can start to sketch out the theory in full. Due to persecution, Jews were pushed into cognitively-demanding occupations like banker or merchant and forced to sink or swim. The ones who swam – people who were intellectually up to the challenge – had more kids than the ones who sank, producing an evolutionary pressure in favor of intelligence greater than that in any other ethnic group. Just as Africans experiencing evolutionary pressure for malaria resistance developed the sickle cell gene, so Ashkenazim experiencing evolutionary pressure for intelligence developed a bunch of genes which increased heterozygotes' IQ but caused serious genetic disease in homozygotes. As a result, Ashkenazi ended up somewhat more intelligent – and somewhat more prone to genetic disease – than the rest of the European population.

If true, this would explain the 27% of Nobel Prizes and 50% of world chess champions thing. But one still has to ask – everywhere had Jews. Why Hungary in particular? What was so special about Budapest in the early 1900s?

IV

Okay, sure, everywhere had Jews. But it's surprising exactly how *many* Jews were in early 1900s Hungary.

The modern United States is about 2% Jewish. Hungary in 1900 was about 5%. The most Jewish city in America, New York, is about 15% Jewish. Budapest in 1900 was 25%. It was one of the most Jewish large cities anywhere in history, excepting only Israel itself. According to Wikipedia, the city's late 19th-century nickname was "Judapest".

So is it possible that all the Jews were winning Nobel Prizes, and Hungary just had more Jews and so more Nobelists?

No. This doesn't seem right. The [1933 European Jewish Population By Country](#) site lists the following size for each country's Jewish communities:

Country	Jewish Population
Poland	3,000,000
Russia	2,500,000
Romania	750,000
Germany	500,000
Hungary	500,000
Britain	300,000
France	250,000
Austria	200,000

It's hard to find a good list of all famous Manhattan Project physicists, but I tried [this article](#) and got the following number of famous Jewish Manhattan Project physicists per country of origin:

Country	MP Physicists
Hungary	4
Germany	2
Poland	2
Austria	2
Italy	1
Netherlands	1

Here's an alternative source with a different definition of "famous", broken down the same way:

Country	MP Physicists
Germany	5
Hungary	4
Poland	3
Italy	2
Austria	2

The main point seems to be disproportionately many people from Central European countries like Hungary and Germany, compared to either Eastern European countries like Poland and Russia or Western European countries like France and Britain.

The Central European advantage over Western Europe is unsurprising; the Western European Jews probably weren't Ashkenazim, and so didn't have the advantage mentioned in the CHH paper above. But is there any reason to think that Central European Jews were more intelligent than Polish and Russian Jews?

I'm not really sure what to think about this. [This paper](#) finds that the sphingolipidoses and other Jewish genetic diseases are about twice as common in Central European Jews as in Eastern European Jews, but I have very low confidence in these results. Intra-Jewish gossip points out the Lithuanians as the geniuses among world Jewry, but doesn't have any similar suggestions about Hungarians. And torsion dystonia, maybe the most clearly IQ-linked disease, is unique to Lithuanians and absent in Hungarians.

Probably much more promising is just to focus on the obvious facts of the social situation. Early 1900s Hungary was a great nation and a prosperous center of learning. Remember, we're talking about the age of the Austro-Hungarian Empire, one of the most industrialized and dynamic economies of the time. It might have had advantages that Poland, Romania, and Russia didn't. My [list of historical national GDPs per capita](#) is very unimpressed by the difference between Hungarian and Polish GDPs in 1900, but maybe it's wrong, or maybe Budapest was an especially modern part of Hungary, or maybe there's something else I'm missing.

Also, there could have been a difference in the position of Jews in these countries. Russia was still experiencing frequent anti-Jewish pogroms in 1900; in Hungary, Jews were among the country's most noble families. Actually, the extent of Jewish wealth and influence in Hungary sort of defies belief. According to Wikipedia, in 1920 Jews were 60% of Hungarian doctors, 50% of lawyers, 40% of engineers and chemists, and 90% of currency brokers and stock exchange members. "In interwar Hungary, more than half and perhaps as much as 90 percent of Hungarian industry was owned or operated by a few closely related Jewish banking families."

So Central European Jews – the Jews in Hungary and Germany – had a unique combination of intellectual and financial advantages. This means Hungary's only real rival here is Germany. Since they were rich, industrialized, and pretty liberal about Jewish rights at the beginning of the 20th century – and since they had just as many Jews as Hungary – we should expect to see the same phenomenon there too.

And we kind of do. Germany produced its share of Jewish geniuses. [Hans Bethe](#) worked for the Manhattan Project and won a Nobel Prize. [Max Born](#) helped develop quantum mechanics and also won a Nobel Prize. [James Franck](#), more quantum physics, another Nobel Prize. [Otto Stern](#), even *more* quantum physics, yet *another* Nobel Prize. [John Polanyi](#), chemical kinetics, Nobel Prize (although he was half-Hungarian). And of course we probably shouldn't forget about

that [Einstein](#) guy. All of these people were born in the same 1880 – 1920 window as the Martians in Hungary.

I think what's going on is this: Germany and Hungary had about the same Jewish population. And they produced about the same number of genius physicists in the same window. But we think of Germany as a big rich country, and Hungary as a small poor country. And the German Jews were spread over a bunch of different cities, whereas the Hungarian Jews were all crammed into Budapest. So when we hear "there were X Nobel Prize winning German physicists in the early 1900s", it sounds only mildly impressive. But when we hear "there were X Nobel Prize winning physicists from Budapest in the early 1900s", it sounds kind of shocking. But the denominator isn't the number of Germans vs. Hungarians, it's the number of German Jews vs. Hungarian Jews, which is about the same.

V

This still leaves one question: why the period 1880 to 1920?

On further reflection, this isn't much of a mystery. The emancipation of the Jews in Eastern Europe was a difficult process that took place throughout the 19th century. Even when it happened, it took a while for the first generation of Jews to get rich enough that their children could afford to go to fancy schools and fritter away their lives on impractical subjects like physics and chemistry. In much of Eastern Europe, the Jews born around 1880 were the first generation that was free to pursue what they wanted and seek their own lot in the world.

The end date around 1920 is more depressing: any Jew born after this time probably wasn't old enough to escape the Nazis. Almost all the famous Hungarian Jews became physics professors in Europe, fled to America during WWII using channels open to famous physicists, and then made most of their achievements on this side of the Atlantic. There are a couple of stragglers born after 1920 who survived – George Soros' family lived because they bought identity documents saying they were Christian; Andrew Grove lived because he was hidden by [righteous Gentiles](#). But in general Jews born in Europe after 1920 didn't have a great life expectancy.

All of this suggests a pretty reasonable explanation of the Martian phenomenon. For the reasons suggested by Cochran, Hardy, and Harpending, Ashkenazi Jews had the potential for very high intelligence. They were mostly too poor and discriminated against to take advantage of it. Around 1880, this changed in a few advanced Central European economies like Germany, Austria, and Hungary. Austria didn't have many Jews. Germany had a lot of Jews, but it was a big country, so nobody really noticed. Hungary had a lot of Jews, all concentrated in Budapest, and so it was really surprising when all of a sudden everyone from Budapest started winning Nobel Prizes around the same time. This continued until World War II, and then all anyone remembered was "Hey, wasn't it funny that so many smart people were born in Budapest between 1880 and 1920?"

And this story is really, really, gloomy.

For centuries, Europe was sitting on this vast untapped resource of potential geniuses. Around 1880, in a few countries only, economic and political conditions finally became ripe for the potential to be realized. The result was one of the greatest spurts of progress in scientific history, bringing us relativity, quantum mechanics, nuclear bombs, dazzling new mathematical systems, the foundations of digital computing, and various other abstruse ideas I don't even pretend to understand. This lasted for approximately one generation, after which a psychopath with a stupid mustache killed everyone involved.

I certainly can't claim that the Jews were the only people being crazy smart in Central Europe around this time. This was the age of Bohr, Schrodinger, Planck, Curie, etc. But part of me wonders even here. If you have one physicist in a town, he sits in an armchair and thinks. If you have five physicists in a town, they meet and talk and try to help each other with their theories. If you have fifty physicists in a town, they can get funding and start a university department.

If you have a hundred, maybe some of them can go into teaching or administration and help support the others. Having this extra concentration of talent in central Europe during this period might have helped Jews and Gentiles alike.

I wonder about this because of a sentiment I hear a lot, from people who know more about physics than I do, that we just don't get people like John von Neumann or Leo Szilard anymore. That there was some weird magical productivity to the early 20th century, especially in Central Europe and Central European immigrants to the United States, that we're no longer really able to match. This can't be a pure numbers game – the Ashkenazi population has mostly recovered since the Holocaust, and people from all over the world are coming to American and European universities and providing more of a concentration of talent than ever. And even though it's impossible to measure, there's still a feeling that it's not enough.

I started down this particular research rabbit hole because a friend challenged me to explain what was so magical about early 20th century Hungary. I think the Jewish population calculations above explain a lot of the story. I'm not sure whether there's a missing ingredient, or, if so, what it might be. Maybe it really was better education. Maybe it really was math competitions and talent searches.

Or maybe it was superintelligent Martian scouts with an Earthling fetish.

Hungarian Education II: Four Nobel Truths

Posted on May 29, 2017 by Scott Alexander

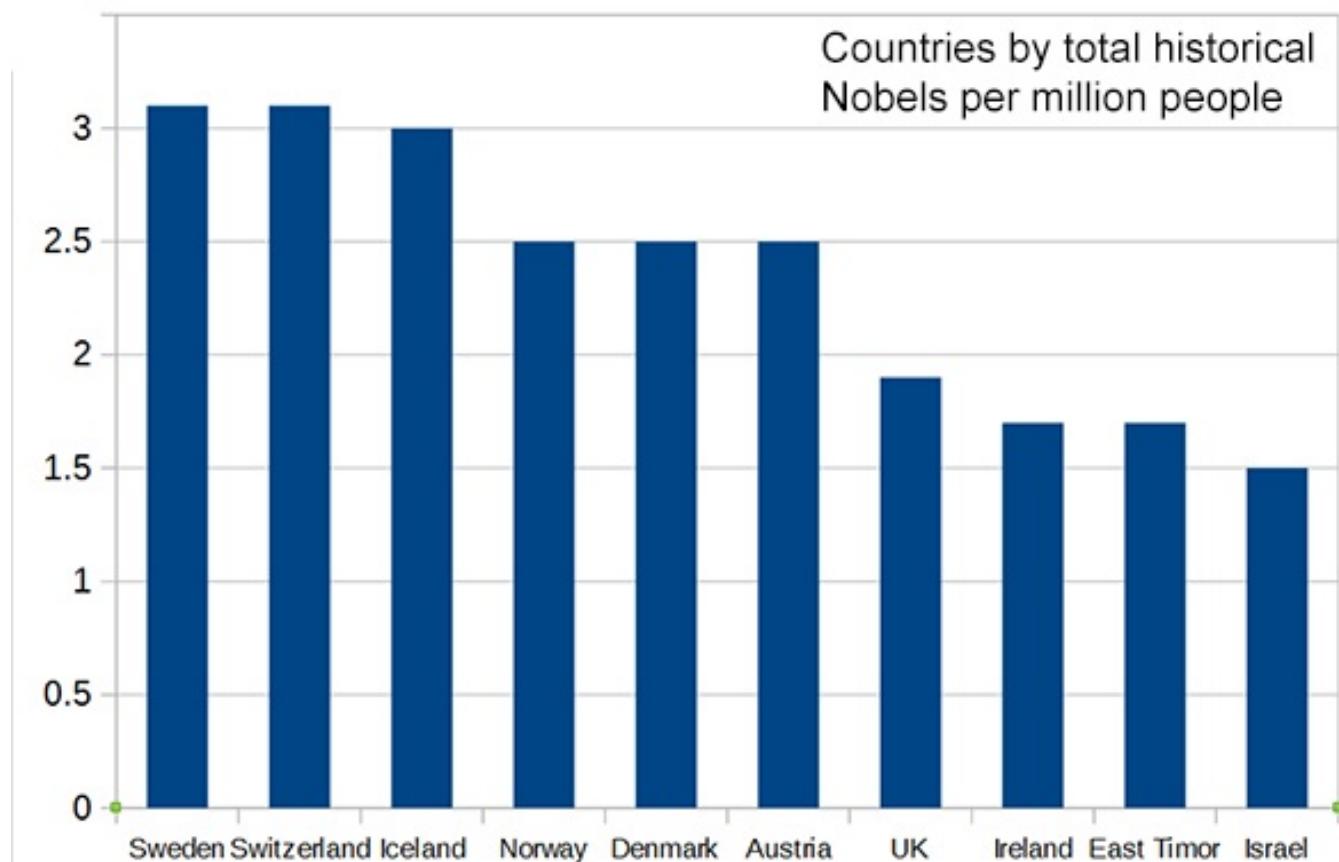


1. Israel historically has only a moderate number of Nobels per capita

On Friday, I discussed [the phenomenon of Hungarian science geniuses](#), and conjectured it was because of Hungary's high concentration of Ashkenazi Jews. A commenter pointed out that Israel had an even higher concentration of Ashkenazi Jews, with less impressive results:

By this logic, Israel should have become the hotbed of geniuses. And while it's true that there are a lot of smart people there, none of the Israeli universities are in the top 10 or maybe even in top 100. And the fraction of Nobel prize winners is not impressive, either.

The objection makes superficial sense. The [list of Nobel winners per capita](#) puts Israel at a modest tenth place, after places like Norway and the United Kingdom.

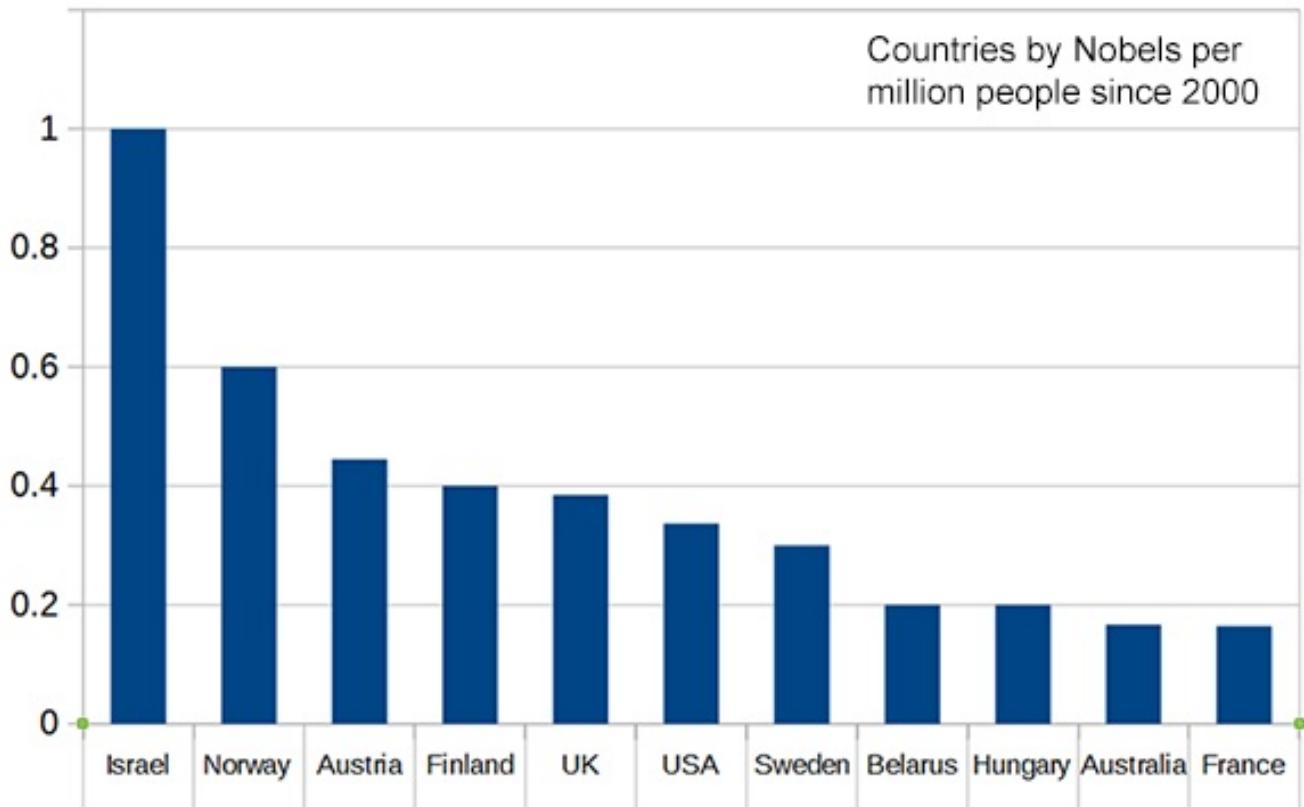


This doesn't look promising for any Ashkenazi-Jew-based theory.

2. But their more modern numbers look much more impressive

On the other hand, that list counts total Nobels won, ever, and divides them by modern population. That gives an advantage to older countries. Norway's been collecting Nobels since 1903; Israel wasn't even *founded* until 1948. And for the first couple generations the Israelis were pretty busy starting kibbutzim, building infrastructure, fighting off enemies, et cetera. Setting up a good university system capable of churning out Nobels takes time. So Norway and the UK had an unfair head start.

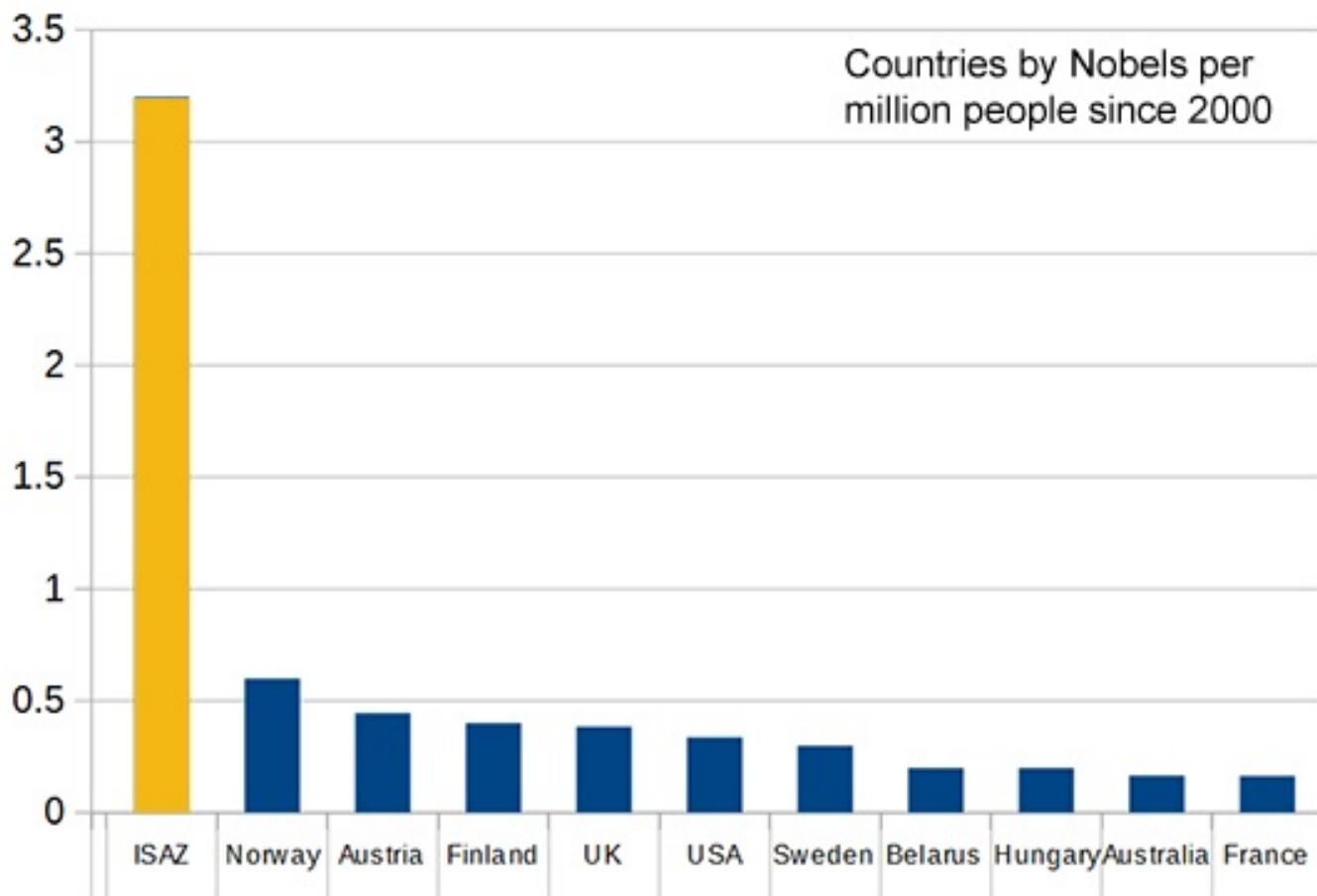
I redid their analysis looking only at Nobels won since the year 2000 (because it was big and round and serves as a signal that I'm trying to avoid optional stopping). My source was [this list of Nobel laureates by country](#), and I deferred to Wikipedia's judgment about whether or not to count dual citizens, immigrants, et cetera. Here's the results:



We see that during this period, Israel has by far the highest number of Nobel prizes per capita.

3. This advantage increases if we look only at Ashkenazim

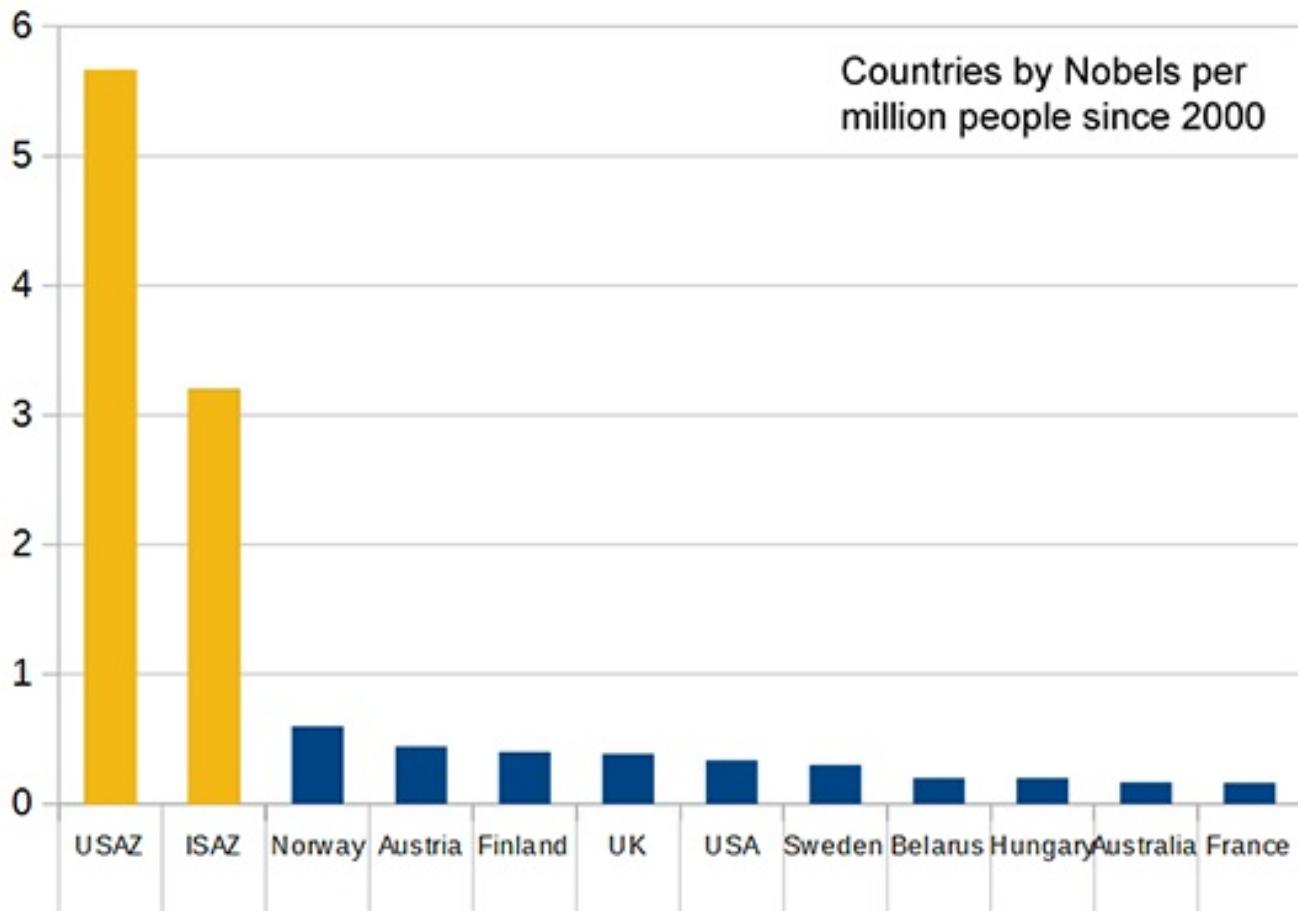
The original theory was about Ashkenazi Jews in particular. Only about a third of Israelis are Ashkenazi (the rest are other types of Jews, or Palestinians, or other non-Jewish minorities). If we separate out the Ashkenazim, the graph looks like this.



ISAZ is Israeli Ashkenazi Jews, considered as a separate population. On the one hand, it's kind of unfair comparing Israel's most successful population group to other countries taken as a whole. On the other hand, if we were to take other countries' most successful population groups, those would be Ashkenazi Jews too, so whatever. Since Israeli Ashkenazim get about five times more Nobels per capita than any country, I'm going to consider the "what about Israel?" objection officially refuted.

4. But there's not a lot of evidence for benefits to concentration, and other factors might be involved

One more graph:



USAZ is US Ashkenazi Jews, who get twice as many Nobels per capita as their Israeli cousins (I'm not sure how seriously to take this; the Israeli data is based on eight Nobel laureates, so there's a lot of room for sampling issues.)

And although it's hard for me to get exact numbers, it looks like a lot of Israeli Nobelists (maybe more than half) did their best work abroad, usually in the US.

Israel went from 1948 to 2002 without winning a single science Nobel (it did win in Literature and Peace during that time). Now it's winning more of them – a lot more, more than any other country per capita – but mostly when its citizens go and study in foreign universities. This seems consistent with an Israeli educational system that's still struggling to get its act together.

Does this mean that once the educational system gets its act together more fully, the ISAZ Nobel rate will approximately double to match the USAZ Nobel rate? I'm not sure.

Just from this analysis, it doesn't look like the theory in the last post, where everyone gets benefits from concentrating closer together, is true. Israel has about ten times as many Ashkenazi Jews per capita than the US, but still does worse than they do.

These data don't challenge the conclusion from the last post that Ashkenazim might have been responsible for Hungary's sudden crop of great scientists. But they do potentially challenge the implicit conclusion that the education system didn't matter that much. I'll have more on that later this week.

Hungarian Education III: Mastering The Core Teachings Of The Budapestians

Posted on May 30, 2017 by Scott Alexander



I

Someone summed up my [previous post](#) as “Hungarian education isn’t magic”. I would amend that to read “Hungarian education isn’t *systemically* magic”. As far as I know, there’s only one Hungarian educator with magic powers, and (like all good wizards) his secrets are maddeningly hard to find.

Laszlo Polgár studied intelligence in university, and decided he had discovered the basic principles behind raising any child to be a genius. He wrote a book called *Bring Up Genius* and recruited an interested woman to marry him so they could test his philosophy by raising children together. He said a bunch of stuff on how ‘natural talent’ was meaningless and so any child could become a prodigy with the right upbringing.

This is normally the point where I’d start making fun of him. Except that when he trained his three daughters in chess, they became the 1st, 2nd, and 6th best female chess players in the world, gaining honors like “youngest grandmaster ever” and “greatest female chess player of all time”. Also they spoke seven languages, including Esperanto.

Their immense success suggests that education can have a major effect even on such traditional genius-requiring domains as chess ability. How can we reconcile that with the rest of our picture of the world, and how obsessed should we be with getting a copy of Laszlo Polgár’s book?

II

Let’s get this out of the way first: the Polgár sisters were probably genetically really smart. The whole family was Hungarian Jews, a group with a great track record. Their mother and father were both well-educated teachers interested in stuff like developmental psychology. They had every possible biological advantage and I’m sure that helped.

J Levitt [proposes](#) an equation to estimate a chess player’s IQ from their chess score. It suggests that chess grandmasters probably have IQs above 160. Plugging the Polgár sisters’ chess scores into his equation, I get IQs in the range of 150, 160, and 170 for the three sisters.

This is biologically impossible. Even if both Polgár parents were 170-IQ themselves, regression to the mean predicts that their children would have IQs around 140 to 150. It’s mathematically possible for there to be an IQ that predicts you would have three children of 150, 160, and 170, but I doubt any living people have it, and even if they did there’s no way they would marry somebody else equally gifted.

EDIT: Thanks to a few people who pointed out some problems with my math here ([1](#), [2](#), [3](#)). I still think that having three supergenius-IQ kids when you and your spouse show no signs of being a supergenius yourself (Laszlo Polgár’s daughters could beat him at chess by the time they were 8) is pretty unlikely, but I admit not impossible. I still think arguing about

this is unnecessary thanks to the points below.

On the other hand, I'm not sure Levitt's right. Chess champion Gary Kasparov actually sat and [took an IQ test](#) for the magazine *Der Spiegel*, and his IQ was 135. That's not bad – it's top 1% of the population – but it's not amazing either.

This is what we should expect given [the correlation of about \$r = 0.24\$](#) between IQ and chess ability (see also [this analysis](#), although I disagree with the details). And the contrary claims – like [the one](#) that Bobby Fischer's IQ was in the 180s – are less well-sourced (although Fischer was [the son of a Hungarian-Jewish mathematician](#), so who knows?).

If it were possible to be a chess world champion with an IQ of 135, then maybe it's possible to be a "mere" grandmaster with IQs in the high 120s and low 130s. And it's just barely plausible that some sufficiently smart people might have three kids who all have IQs in the high 120s and low 130s.

But this just passes the buck on the mystery. 2% of people have IQs in the high 120s or low 130s, but 2% of people aren't the top-ranked female chess player in the world. The Polgár sisters' IQs might have been a *permissive* factor in allowing them to excel, but it didn't *necessitate* it. So what's going on there?

III

"Practice" seems like an obvious part of the picture. Malcolm Gladwell uses the Polgárs as poster children for his famous '[10,000 hours of practice makes you an expert at anything](#)' rule. The Polgárs had 50,000 hours of chess practice each by the time they were adults, presumably enough to make them quintuple-experts.

Robert Howard has a paper [Does High-Level Performance Depend On Practice Alone? Debunking The Polgár Sisters Case](#) in which he argues against the strong version of Gladwell's thesis. He points out that there are many chess masters who have practiced much less than the Polgár sisters but are better than they are. He also points out that even though the sisters themselves have all practiced similar amounts, youngest sister Judit is clearly better than the other two in a way that practice alone cannot explain.

I don't know if the case he's arguing against – that practice is literally everything and it's impossible for anything else to factor in – is a straw man or not. But it seems more important to consider a less silly argument – that practice is one of many factors, and that enough of it can make up for a lack of the others. This seems potentially true. [This study](#) showing that amount of practice only explains 12% of the variance in skill level at various tasks, and is often summarized as "practice doesn't matter much". But it finds practice matters more (25% of the variance) in unchanging games with clear fixed rules, and uses chess as an example.

So suppose that the Polgár sisters are genetically smart, but maybe not as high up there as some other chess masters. We would expect them to need much more practice to achieve a level of proficiency similar to those chess masters, and indeed that seems like what happens.

(all of this is confounded by them being women and almost all the other equally-good chess masters being men. It's unclear if the Polgárs deserve extra points for overcoming whatever factor usually keeps women out of the highest levels of chess.)

But I'm actually still not sure this suffices as an explanation. According to Wikipedia:

Polgár began teaching his eldest daughter, Susan, to play chess when she was four years old. Six months later, Susan toddled into Budapest's smoke-filled chess club," which was crowded with elderly men, and

proceeded to beat the veteran players.

The study linked above suggests that Susan practiced 48 hours a week. During those six months, she would have accumulated about 1200 hours of practice. Suppose the elderly Budapest chess players practiced only one hour a week, but had been doing so for the last twenty-five years. They would have more practice than Susan – plus the advantage of having older, more developed brains. So why did she beat them so easily?

Maybe there's a time-decay factor for practice? That is, maybe Susan had been practicing intensively, so she got a lot of chances to link it all together as she was learning, and also it was fresh in her mind when she went to the club to go play? I'm not sure. If some of those veterans had been playing more than one hour a week (and surely the sort of people who frequent Budapest chess clubs do) then her advantage seems too implausible to be due to freshness-of-material alone.

IV

That leaves two possibilities.

First, Susan could have benefitted from some form of malleability. A lot of people claim there's a "developmental window" during which children have a unique ability to learn language. [If cats see only vertical stripes](#) for the first few weeks of their lives, they never learn to see in horizontal. Maybe if you teach your kid high-level chess at age 4, they'll be able to recruit systems that adults could never manage, or reorganize the fundamental structure of their brain to conform to chess better, or something like that.

Second, Polgár might actually have some really good educational methods besides just "start early and have a lot of practice". I assume this is true, but I'm having a lot of trouble finding them. Shockingly, Polgár's book *Bring Up Genius* is out of print and totally unavailable anywhere – I guess the book-reading community heard that someone wrote down a way to reliably turn any child into a genius which had a great real-world track record of success, and collectively decided "Nah, better read *Fifty Shades Of Grey* instead". I'm not sure at what point I should start positing a conspiracy of suppression, or whether that would be better or worse than the alternative.

The book seems to possibly be available in Hungarian under the title *Nevelj zsenit!*, but I can't tell for sure and a lot of the Hungarian sites suggest it's out of print even in that language. There *may* have been a recent republication in Esperanto called *Eduku genulon!*, but I can't find that one either. If anybody knows where to find this book and wants to send it to me, I will figure out some way to translate it and review it. I'd also be willing to pay for costs and even pay extra for your time if it helps. Come on, Esperanto-speakers! This is the only chance you'll ever have to be useful!

V

One thing I know without reading the book: Polgár says that his method should work to create geniuses in any field, not just chess. He said he chose chess kind of on the whim of his eldest daughter. From Wikipedia:

Polgár and his wife considered various possible subjects in which to drill their children, "including mathematics and foreign languages," but they settled on chess. "We could do the same thing with any subject, if you start early, spend lots of time and give great love to that one subject," Klara later explained. "But we chose chess. Chess is very objective and easy to measure." Susan described chess as having been her own choice: "Yes, he could have put us in any field, but it was I who chose chess as a four-year-old.... I liked the chessmen; they were toys for me."

It's disappointing that he decided to stick with chess for his other two daughters. The study linked above suggests that chess is unusually amenable to practice. What would have happened if he'd tried to train his kids in art? In mathematics? In entrepreneurship? I'm not sure, and I'm really tempted to have some kids and find out.

(be right back, going to change my OkCupid profile to include "must be interested in n=1 developmental-psych experiments, have access to a rare book library, and speak either Hungarian or Esperanto")

I mentioned this plan to a friend, who protested that this was cruel and tantamount to child abuse. After all, how can you force someone to spend their entire childhood indoors, studying mind-numbing chess problems day in and day out, instead of enjoying themselves like normal kids?

First of all, this isn't how the Polgár children (or adults) describe their experiment. From [The Guardian](#):

Starting with his eldest daughter, Susan, Polgár was careful to treat it as a playful activity, turning it into a fantasy of dramatic wins and losses. Whereas Earl and Kultida Woods had coerced perfection from Tiger, the Polgárs encouraged enjoyment. By the time Susan had turned five, she was excited by playing and spent hundreds of hours practising. She was entered into a local competition and treated it as fun, winning 10-0, causing a sensation.

Meanwhile, her younger sisters were intrigued and László allowed them to feel the pieces, seeing them as toys, with no formal tuition until they were five. Interviewed recently, all three girls described playing the game as something that they loved doing – it never felt like a chore. Instead of messing about playing Monopoly, netball or going to the local swimming pool, chess was just what the Polgár family enjoyed... Polgár understood that coercion was less valuable than small children's need to enjoy fantasy play. Consequently, his daughters all seem to have grown into satiable, well-balanced people rather than success addicts.

But more important – I responded that the Polgárs claim to have spent about 48 hours a week practicing chess. I spent seven hours a day in school, so if my teachers assigned two hours of homework a night then we spent about the same amount of time getting educated. Except what the Polgárs got out of it was world-champion-level mastery of their favorite subject in the world, nationwide fame, and (by their own accounts) loving every second of it, and what I got was staring out a window all day as my teacher declared that we were going to make a collage about the meaning of Respect.

The Polgár sisters talk about how they loved their education, had a great childhood, thought their parents were always patient with them and never strict and harsh, and don't regret anything. How many kids who went to public school can say the same?

An article about Laszlo Polgár mentions that he had to fight the Hungarian authorities to be allowed to home school his children. Imagine being so certain of your own home-schooling techniques that you're afraid taking your kids to [the Fasori Gymnasium](#) is going to stunt their intellectual growth. And imagine being *right*. And imagine my friend thinking that normal American public school might be better than that. It sort of boggles the imagination.

And I guess I shouldn't be too harsh, because the public school system tries to do the best it can with an impossible set of constraints. But I'm still suspicious. *Who else has the motivation to hide that book?*

EDIT: Thanks to readers, I've got an Esperanto copy and a person willing to translate it. I'll let you know as this develops.

Book Review: Seeing Like A State

Posted on March 16, 2017 by Scott Alexander



I

[Seeing Like A State](#) is the book G.K. Chesterton would have written if he had gone into economic history instead of literature. Since he didn't, James Scott had to write it a century later. The wait was worth it.

Scott starts with the story of "scientific forestry" in 18th century Prussia. Enlightenment rationalists noticed that peasants were just cutting down whatever trees happened to grow in the forests, *like a chump*. They came up with a better idea: clear all the forests and replace them by planting identical copies of Norway spruce (the highest-lumber-yield-per-unit-time tree) in an evenly-spaced rectangular grid. Then you could just walk in with an axe one day and chop down like a zillion trees an hour and have more timber than you could possibly ever want.

This went poorly. The impoverished ecosystem couldn't support the game animals and medicinal herbs that sustained the surrounding peasant villages, and they suffered an economic collapse. The endless rows of identical trees were a perfect breeding ground for plant diseases and forest fires. And the complex ecological processes that sustained the soil stopped working, so after a generation the Norway spruces grew stunted and malnourished. Yet for some reason, everyone involved got promoted, and "scientific forestry" spread across Europe and the world.

And this pattern repeats with suspicious regularity across history, not just in biological systems but also in social ones.

Natural organically-evolved cities tend to be densely-packed mixtures of dark alleys, tiny shops, and overcrowded streets. Modern scientific rationalists came up with a better idea: an evenly-spaced rectangular grid of identical giant Brutalist apartment buildings separated by wide boulevards, with everything separated into carefully-zoned districts. Yet for some reason, whenever these new rational cities were built, people hated them and did everything they could to move out into more organic suburbs. And again, for some reason the urban planners got promoted, became famous, and spread their destructive techniques around the world.

Ye olde organically-evolved peasant villages tended to be complicated confusions of everybody trying to raise fifty different crops at the same time on awkwardly shaped cramped parcels of land. Modern scientific rationalists came up with a better idea: giant collective mechanized farms growing purpose-bred high-yield crops and arranged in (say it with me) evenly-spaced rectangular grids. Yet for some reason, these giant collective farms had lower yields per acre than the old traditional methods, and wherever they arose famine and mass starvation followed. And again, for some reason governments continued to push the more "modern" methods, whether it was socialist collectives in the USSR, big agricultural corporations in the US, or sprawling banana plantations in the Third World.

Traditional lifestyles of many East African natives were nomadic, involving slash-and-burn agriculture in complicated jungle terrain according to a bewildering variety of ad-hoc rules. Modern scientific rationalists in African governments (both colonial and independent) came up with a better idea – resettlement of the natives into villages, where they could have modern amenities like schools, wells, electricity, and evenly-spaced rectangular grids. Yet for some reason, these villages kept failing: their crops died, their economies collapsed, and their native inhabitants disappeared back into the jungle. And again, for some reason the African governments kept trying to bring the natives back and make them stay, even if they had to blur the lines between villages and concentration camps to make it work.



A favorite Seeing Like A State image: a comparison of street maps for Bruges (a premodern organic city) with Chicago (a modern planned city)

Why did all of these schemes fail? And more importantly, why were they celebrated, rewarded, and continued, even when the fact of their failure became too obvious to ignore? Scott gives a two part answer.

The first part of the story is High Modernism, an aesthetic taste masquerading as a scientific philosophy. The High Modernists claimed to be about figuring out the most efficient and high-tech way of doing things, but most of them knew little relevant math or science and were basically just LARPing being rational by placing things in evenly-spaced rectangular grids.

But the High Modernists were pawns in service of a deeper motive: the centralized state wanted the world to be “legible”, ie arranged in a way that made it easy to monitor and control. An intact forest might be more productive than an evenly-spaced rectangular grid of Norway spruce, but it was harder to legislate rules for, or assess taxes on.

The state promoted the High Modernists’ platitudes about The Greater Good as cover, in order to implement the totalitarian schemes they wanted to implement anyway. The resulting experiments were usually failures by the humanitarian goals of the Modernists, but resounding successes by the command-and-control goals of the state. And so we gradually transitioned from systems that were messy but full of fine-tuned hidden order, to ones that were barely-functional but really easy to tax.

II

Suppose you’re a premodern king, maybe one of the Louises who ruled France in the Middle Ages. You want to tax people to raise money for a Crusade or something. Practically everyone in your kingdom is a peasant, and all the peasants produce is grain, so you’ll tax them in grain. Shouldn’t be too hard, right? You’ll just measure how many pints of grain everyone produces, and...

The pint in eighteenth-century Paris was equivalent to 0.93 liters, whereas in Seine-en-Montane it was 1.99 liters and in Precy-sous-Thil, an astounding 3.33 liters. The aune, a measure of length used for cloth, varied depending on the material (the unit for silk, for instance, was smaller than that for linen) and across France there were at least seventeen different aunes.

Okay, this is stupid. Just give everybody evenly-sized baskets, and tell them that baskets are the new unit of measurement.

Virtually everywhere in early modern Europe were endless micropolitics about how baskets might be adjusted through wear, bulging, tricks of weaving, moisture, the thickness of the rim, and so on. In some areas the local standards for the bushel and other units of measurement were kept in metallic form and placed in the care of a trusted official or else literally carved into the stone of a church or the town hall. Nor did it end there. How the grain was to be poured (from shoulder height, which packed it somewhat, or from waist height?), how damp it could be, whether the container could be shaken down, and finally, if and how it was to be leveled off when full were subjects of long and bitter controversy.

Huh, this medieval king business is harder than you thought. Maybe you can just leave this problem to the feudal lords?

Thus far, this account of local measurement practices risks giving the impression that, although local conceptions of distance, area, volume, and so on were different from and more varied than the unitary abstract standards a state might favor, they were nevertheless aiming at objective accuracy. This impression would be false. [...]

A good part of the politics of measurement sprang from what a contemporary economist might call the “stickiness” of feudal rents. Noble and clerical claimants often found it difficult to increase feudal dues directly; the levels set for various charges were the result of long struggle, and even a small increase above the customary level was viewed as a threatening breach of tradition. Adjusting the measure, however, represented a roundabout way of achieving the same end.

The local lord might, for example, lend grain to peasants in smaller baskets and insist on repayment in larger baskets. He might surreptitiously or even boldly enlarge the size of the grain sacks accepted for milling (a monopoly of the domain lord) and reduce the size of the sacks used for measuring out flour; he might also collect feudal dues in larger baskets and pay wages in kind in smaller baskets. While the formal custom governing feudal dues and wages would thus remain intact (requiring, for example, the same number of sacks of wheat from the harvest of a given holding), the actual transaction might increasingly favor the lord. The results of such fiddling were far from trivial. Kula estimates that the size of the bushel (boisseau) used to collect the main feudal rent (*taille*) increased by one-third between 1674 and 1716 as part of what was called the reaction feodale.

Okay, but nobody's going to make *too* big a deal about this, right?

This sense of victimization [over changing units of measure] was evident in the cahiers of grievances prepared for the meeting of the Estates General just before the Revolution. [...] In an unprecedented revolutionary context where an entirely new political system was being created from first principles, it was surely no great matter to legislate uniform weights and measures. As the revolutionary decree read “The centuries old dream of the masses of only one just measure has come true! The Revolution has given the people the meter!”

Okay, so apparently (you think to yourself as you are being led to the guillotine), it was a big deal after all.

Maybe you *shouldn't* have taxed grain. Maybe you should tax land. After all, it's the land that grows the grain. Just figure out how much land everybody owns, and you can calculate some kind of appropriate tax from there.

So, uh, peasant villagers, how much land does each of you own?

A hypothetical case of customary land tenure practices may help demonstrate how difficult it is to assimilate such practices to the barebones scheme of a modern cadastral map [land survey suitable for tax assessment] [...]

Let us imagine a community in which families have usufruct rights to parcels of cropland during the main growing season. Only certain crops, however, may be planted, and every seven years the usufruct land is distributed among resident families according to each family's size and its number of able-bodied adults. After the harvest of the main-season crop, all cropland reverts to common land where any family may glean, graze their fowl and livestock, and even plant quickly maturing, dry-season crops. Rights to graze fowl and livestock on pasture-land held in common by the village is extended to all local families, but the number of animals that can be grazed is restricted according to family size, especially in dry years when forage is scarce. Families not using their grazing rights can give them to other villagers but not to outsiders. Everyone has the right to gather firewood for normal family needs, and the village blacksmith and baker are given larger allotments. No commercial sale from village woodlands is permitted.

Trees that have been planted and any fruit they may bear are the property of the family who planted them, no matter where they are now growing. Fruit fallen from such tree, however, is the property of anyone who gathers it. When a family fells one of its trees or a tree is felled by a storm, the trunk belongs to the family, the branches to the immediate neighbors, and the "tops" (leaves and twigs) to any poorer villager who carries them off. Land is set aside for use or leasing out by widows with children and dependents of conscripted males. Usufruct rights to land and trees may be let to anyone in the village; the only time they may be let to someone outside the village is if no one in the community wishes to claim them. After a crop failure leading to a food shortage, many of these arrangements are readjusted.

You know what? I'm just going to put you all down as owning ten. Ten land. Everyone okay with that? Cool. Let's say ten land for everyone and just move on to the next village.

Novoselok village had a varied economy of cultivation, grazing, and forestry... the complex welter of strips was designed to ensure that each village household received a strip of land in every ecological zone. An individual household might have as many as ten to fifteen different plots constituting something of a representative sample of the village's ecological zones and microclimates. The distribution spread a family's risks prudently, and from time to time the land was reshuffled as families grew or shrunk... The strips of land were generally straight and parallel so that a readjustment could be made by moving small stakes along just one side of a field, without having to think of areal dimensions. Where the other side of the field was not parallel, the stakes could be shifted to compensate for the fact that the strip lay toward the narrower or wider end of the field. Irregular fields were divided, not according to area, but according to yield.

...huh. Maybe this isn't going to work. Let's try it the other way around. Instead of mapping land, we can just get a list with the name of everyone in the village, and go from there.

Only wealthy aristocrats tended to have fixed surnames... Imagine the dilemma of a tithe or capitation-tax collector [in England] faced with a male population, 90% of whom bore just six Christian names (John, William, Thomas, Robert, Richard, and Henry).

Okay, fine. That won't work either. Surely there's *something* else we can do to assess a tax burden on each estate.

Think outside the box, scrape the bottom of the barrel!

The door-and-window tax established in France [in the 18th century] is a striking case in point. Its originator must have reasoned that the number of windows and doors in a dwelling was proportional to the dwelling's size. Thus a tax assessor need not enter the house or measure it, but merely count the doors and windows.

As a simple, workable formula, it was a brilliant stroke, but it was not without consequences. Peasant dwellings were subsequently designed or renovated with the formula in mind so as to have as few openings as possible. While the fiscal losses could be recouped by raising the tax per opening, the long-term effects on the health of the population lasted for more than a century.

Close enough.

III

The moral of the story is: premodern states had very limited ability to tax their citizens effectively. Along with the problems mentioned above – nonstandardized measurement, nonstandardized property rights, nonstandardized personal names – we can add a few others. At this point national languages were a cruel fiction; local “dialects” could be as different from one another as eg Spanish is from Portuguese, so villagers might not even be able to *understand* the tax collectors. Worst of all, there was no such thing as a census in France until the 17th century, so there wasn’t even a good idea of how many people or villages there were.

Kings usually solved this problem by leaving the tax collection up to local lords, who presumably knew the idiosyncrasies of their own domains. But one step wasn’t always enough. If the King only knew Dukes, and the Dukes only knew Barons, and the Barons only knew village headmen, and it was only the village headmen who actually knew anything about the peasants, then you needed a four-step chain to get any taxes. Each link in the chain had an incentive to collect as much as they could and give up as little as they could get away with. So on the one end, the peasants were paying backbreaking punitive taxes. And on the other, the Royal Treasurer was handing the King half a loaf of moldy bread and saying “Here you go, Sire, apparently this is all the grain in France.”

So from the beginning, kings had an incentive to make the country “legible” – that is, so organized and well-indexed that it was easy to know everything about everyone and collect/double-check taxes. Also from the beginning, nobles had an incentive to frustrate the kings so that they wouldn’t be out of a job. And commoners, who figured that anything which made it easier for the State to tax them and interfere in their affairs was bad news, usually resisted too.

Scott doesn’t bring this up, but it’s interesting reading this in the context of Biblical history. It would seem that whoever wrote the Bible was not a big fan of censuses. From [1 Chronicles 21](#):

Satan rose up against Israel and caused David to take a census of the people of Israel. So David said to Joab and the commanders of the army, “Take a census of all the people of Israel—from Beersheba in the south to Dan in the north—and bring me a report so I may know how many there are.”

But Joab replied, “May the Lord increase the number of his people a hundred times over! But why, my lord the king, do you want to do this? Are they not all your servants? Why must you cause Israel to sin?”

But the king insisted that they take the census, so Joab traveled throughout all Israel to count the people.

Then he returned to Jerusalem and reported the number of people to David. There were 1,100,000 warriors in all Israel who could handle a sword, and 470,000 in Judah. But Joab did not include the tribes of Levi and Benjamin in the census because he was so distressed at what the king had made him do.

God was very displeased with the census, and he punished Israel for it. Then David said to God, "I have sinned greatly by taking this census. Please forgive my guilt for doing this foolish thing." Then the Lord spoke to Gad, David's seer. This was the message: "Go and say to David, 'This is what the Lord says: I will give you three choices. Choose one of these punishments, and I will inflict it on you.'"

So Gad came to David and said, "These are the choices the Lord has given you. You may choose three years of famine, three months of destruction by the sword of your enemies, or three days of severe plague as the angel of the Lord brings devastation throughout the land of Israel. Decide what answer I should give the Lord who sent me."

"I'm in a desperate situation!" David replied to Gad. "But let me fall into the hands of the Lord, for his mercy is very great. Do not let me fall into human hands."

So the Lord sent a plague upon Israel, and 70,000 people died as a result.

(related: Scott examined some of the same data about Holocaust survival rates as *Eichmann In Jerusalem*, but made them make a lot more sense: the greater the legibility of the state, the worse for the Jews. One reason Jewish survival in the Netherlands was so low was because the Netherlands had a very accurate census of how many Jews there were and where they lived; sometimes officials saved Jews by literally burning census records).

Centralized government projects promoting legibility have always been a two-steps-forward, one-step back sort of thing. The government very gradually expands its reach near the capital where its power is strongest, to peasants whom it knows will try to thwart it as soon as its back is turned, and then if its decrees survive it pushes outward toward the hinterlands.

Scott describes the spread of surnames. Peasants didn't like permanent surnames. Their own system was quite reasonable for them: John the baker was John Baker, John the blacksmith was John Smith, John who lived under the hill was John Underhill, John who was really short was John Short. The same person might be John Smith and John Underhill in different contexts, where his status as a blacksmith or place of origin was more important.

But the government insisted on giving everyone a single permanent name, unique for the village, and tracking who was in the same family as whom. Resistance was intense:

What evidence we have suggests that second names of any kind became rare as distance from the state's fiscal reach increased. Whereas one-third of the households in Florence declared a second name, the proportion dropped to one-fifth for secondary towns and to one-tenth in the countryside. It was not until the seventeenth century that family names crystallized in the most remote and poorest areas of Tuscany – the areas that would have had the least contact with officialdom. [...]

State naming practices, like state mapping practices, were inevitably associated with taxes (labor, military service, grain, revenue) and hence aroused popular resistance. The great English peasant rising of 1381 (often called the Wat Tyler Rebellion) is attributed to an unprecedented decade of registration and assessments of poll taxes. For English as well as for Tuscan peasants, a census of all adult males could not but appear ominous, if not ruinous.

The same issues repeated themselves a few hundred years later when Europe started colonizing other continents. Again they encountered a population with naming systems they found unclear and unsuitable to taxation. But since colonial states had more control over their subjects than the relatively weak feudal monarchies of the Middle Ages, they were able to deal with it in one fell swoop, sometimes comically so:

Nowhere is this better illustrated than in the Philippines under the Spanish. Filipinos were instructed by the decree of November 21, 1849 to take on permanent Hispanic surnames. [...]

Each local official was to be given a supply of surnames sufficient for his jurisdiction, “taking care that the distribution be made by letters of the alphabet.” In practice, each town was given a number of pages from the alphabetized [catalog], producing whole towns with surnames beginning with the same letter. In situations where there has been little in-migration in the past 150 years, the traces of this administrative exercise are still perfectly visible across the landscape. “For example, in the Bikol region, the entire alphabet is laid out like a garland over the provinces of Albay, Sorsogon, and Catanduanes which in 1849 belonged to the single jurisdiction of Albay. Beginning with A at the provincial capital, the letters B and C mark the towns along the coast beyond Tabaco to Wiki. We return and trace along the coast of Sorsogon the letters E to L, then starting down the Iraya Valley at Daraga with M, we stop with S to Polangui and Libon, and finish the alphabet with a quick tour around the island of Catanduanes.

The confusion for which the decree is the antidote is largely that of the administrator and the tax collector. Universal last names, they believe, will facilitate the administration of justice, finance, and public order as well as make it simpler for prospective marriage partners to calculate their degree of consanguinity. For a utilitarian state builder of [Governor] Claveria’s temper, however, the ultimate goal was a complete and legible list of subjects and taxpayers.

This was actually a lot less cute and funny than the alphabetization makes it sound:

What if the Filipinos chose to ignore their new last names? This possibility had already crossed Claveria’s mind, and he took steps to make sure that the names would stick. Schoolteachers were ordered to forbid their students to address or even know one another by any name except the officially inscribed family name. Those teachers who did not apply the rule with enthusiasm were to be punished. More efficacious perhaps, given the minuscule school enrollment, was the proviso that forbade priests and military and civil officials from accepting any document, application, petition, or deed that did not use the official surnames. All documents using other names would be null and void

Similar provisions ensured the replacement of local dialects with the approved national language. Students were only allowed to learn the national language in school and were punished for speaking in vernacular. All formal documents had to be in the national language, which meant that peasants who had formally been able to manage their own legal affairs had to rely on national-language-speaking intermediaries. Scott talks about the effect in France:

One can hardly imagine a more effective formula for immediately devaluing local knowledge and privileging all those who had mastered the official linguistic code. It was a gigantic shift in power. Those at the periphery who lacked competence in French were rendered mute and marginal. They were now in need of a local guide to the new state culture, which appeared in the form of lawyers, notaries, schoolteachers, clerks, and soldiers.

IV

So the early modern period is defined by an uneasy truce between states who want to be able to count and standardize everything, and citizens who don't want to let them. Enter High Modernism. Scott defines it as

A strong, one might even say muscle-bound, version of the self-confidence about scientific and technical progress, the expansion of production, the growing satisfaction of human needs, the mastery of nature (including human nature), and above all, the rational design of social order commensurate with the scientific understanding of natural laws

...which is just a *bit* academic-ese for me. An extensional definition might work better: standardization, Henry Ford, the factory as metaphor for the best way to run everything, conquest of nature, New Soviet Man, people with college degrees knowing better than you, wiping away the foolish irrational traditions of the past, Brave New World, everyone living in dormitories and eating exactly 2000 calories of Standardized Food Product (TM) per day, anything that is For Your Own Good, gleaming modernist skyscrapers, The X Of The Future, complaints that the unenlightened masses are resisting The X Of The Future, demands that if the unenlightened masses reject The X Of The Future they must be re-educated For Their Own Good, and (of course) evenly-spaced rectangular grids.

(maybe the best definition would be “everything G. K. Chesterton didn’t like.”)

It sort of sounds like a Young Adult Dystopia, but Scott shocked me with his research into just how strong this ideology was around the turn of the last century. Some of the greatest early 20th-century thinkers were High Modernist to the point of self-parody, the point where a Young Adult Dystopian fiction writer would start worrying they were laying it on a little too thick.

The worst of the worst was Le Corbusier, the French artist/intellectual/architect. The Soviets asked him to come up with a plan to redesign Moscow. He came up with one: kick out everyone, bulldoze the entire city, and redesign it from scratch upon rational principles. For example, instead of using other people’s irrational systems of measurement, they would use a new measurement system invented by Le Corbusier himself, called [Modulor](#), which combined the average height of a Frenchman with the Golden Ratio.



Also, evenly-spaced rectangular grids may have been involved.

The Soviets decided to pass: the plan was too extreme and destructive of existing institutions *even for Stalin*. Undeterred, Le Corbusier changed the word "Moscow" on the diagram to "Paris", then presented it to the French government (who also passed). Some aspects of his design eventually ended up as Chandigarh, India.



A typical building in Chandigarh. The Soviets and French must have been kicking themselves when they realized what they'd missed out on.

Le Corbusier was challenged on his obsession with keeping his plan in the face of different local conditions, pre-existing structures, residents who might want a say in the matter, et cetera. Wasn't it kind of dictatorial? He replied that:

The despot is not a man. It is the Plan. The correct, realistic, exact plan, the one that will provide your solution once the problem has been posited clearly, in its entirety, in its indispensable harmony. This plan has been drawn up well away from the frenzy in the mayor's office or the town hall, from the cries of the electorate or the laments of society's victims. It has been drawn up by serene and lucid minds. It has taken account of nothing but human truths. It has ignored all current regulations, all existing usages, and channels. It has not considered whether or not it could be carried out with the constitution now in force. It is a biological creation destined for human beings and capable of realization by modern techniques.

What was so great about this "biological creation" of "serene and lucid minds"? It... might have kind of maybe been evenly-spaced rectangular grids:

People will say: "That's easily said! But all your intersections are right angles. What about the infinite variations that constitute the reality of our cities?" But that's precisely the point: I eliminate all these things. Otherwise we shall never get anywhere.

I can already hear the storms of protest and the sarcastic gibes: "Imbecile, madman, idiot, braggart, lunatic, etc." Thank you very much, but it makes no difference: my starting point is still the same: I insist on right-angled intersections. The intersections shown here are all perfect.

Scott uses Le Corbusier as the epitome of five High Modernist principles.

First, there can be no compromise with the existing infrastructure. It was designed by superstitious people who didn't have architecture degrees, or at the very least got their architecture degrees in the past and so were insufficiently Modern. The more completely it is bulldozed to make way for the Glorious Future, the better.

Second, human needs can be abstracted and calculated. A human needs X amount of food. A human needs X amount of water. A human needs X amount of light, and prefers to travel at X speed, and wants to live within X miles of the workplace. These needs are easily calculable by experiment, and a good city is the one built to satisfy these needs and ignore any competing frivolities.

Third, the solution is the solution. It is universal. The rational design for Moscow is the same as the rational design for Paris is the same as the rational design for Chandigarh, India. As a corollary, all of these cities ought to look exactly the same. It is *maybe* permissible to adjust for obstacles like mountains or lakes. But only if you are on too short a budget to follow the rationally correct solution of leveling the mountain and draining the lake to make your city truly optimal.

Fourth, all of the relevant rules should be explicitly determined by technocrats, then followed to the letter by their subordinates. Following these rules is better than trying to use your intuition, in the same way that using the laws of physics to calculate the heat from burning something is better than just trying to guess, or following an evidence-based clinical algorithm is better than just prescribing whatever you feel like.

Fifth, there is nothing whatsoever to be gained or learned from the people involved (eg the city's future citizens). *You* are a rational modern scientist with an architecture degree who has already calculated out the precise value for all relevant urban parameters. *They* are yokels who probably cannot even *spell* the word architecture, let alone usefully contribute to it. They probably make all of their decisions based on superstition or tradition or something, and their input should be ignored For Their Own Good.

And lest I be unfair to Le Corbusier, a lot of his scientific rational principles made a lot of sense. Have wide roads so that there's enough room for traffic and all the buildings get a lot of light. Use rectangular grids to make cities easier to navigate. Avoid frivolous decoration so that everything is efficient and affordable to all. Use concrete because it's the cheapest and strongest material. Keep pedestrians off the streets as much as possible so that they don't get hit by cars. Use big apartment towers to save space, then use the open space for pretty parks and public squares. Avoid anything that looks like a local touch, because nationalism leads to war and we are all part of the same global community of humanity. It sounded pretty good, and for a few decades the entire urban planning community was convinced.

So, how did it go?

Scott uses the example of Brasilia. Brazil wanted to develop its near-empty central regions and decided to build a new capital in the middle of nowhere. They hired three students of Le Corbusier, most notably Oscar Niemeyer, to build them a perfect scientific rational city. The conditions couldn't have been better. The land was already pristine, so there was no need to bulldoze Paris first. There were no inconvenient mountains or forests in the way. The available budget was in the tens of billions. The architects rose to the challenge and built them the world's greatest High Modernist city.



It's... even more beautiful than I imagined

Yet twenty years after its construction, the city's capacity of 500,000 residents was only half-full. And it wasn't the location – a belt of suburbs grew up with a population of almost a million. People wanted to live in *the vicinity of* Brasília. They just didn't want to live *in the parts that Niemeyer and the Corbusierites had built*.



Brasília from above. Note both the evenly-spaced rectangular grid of identical buildings in the center, and the fact that most people aren't living in it.

What happened? Scott writes:

Most of those who have moved to Brasília from other cities are amazed to discover "that it is a city without crowds." People complain that Brasília lacks the bustle of street life, that it has none of the busy street

corners and long stretches of storefront facades that animate a sidewalk for pedestrians. For them, it is almost as if the founders of Brasilia, rather than having planned a city, have actually planned to prevent a city. The most common way they put it is to say that Brasilia “lacks street corners,” by which they mean that it lacks the complex intersections of dense neighborhoods comprising residences and public cafes and restaurants with places for leisure, work, and shopping.

While Brasilia provides well for some human needs, the functional separation of work from residence and of both from commerce and entertainment, the great voids between superquadra, and a road system devoted exclusively to motorized traffic make the disappearance of the street corner a foregone conclusion. The plan did eliminate traffic jams; it also eliminated the welcome and familiar pedestrian jams that one of Holston’s informants called “the point of social conviviality

The term brasilité, meaning roughly Brasilia-itis, which was coined by the first-generation residents, nicely captures the trauma they experienced. As a mock clinical condition, it connotes a rejection of the standardization and anonymity of life in Brasilia. “They use the term brasilité to refer to their feelings about a daily life without the pleasures—the distractions, conversations, flirtations, and little rituals of outdoor life in other Brazilian cities.” Meeting someone normally requires seeing them either at their apartment or at work. Even if we allow for the initial simplifying premise of Brasilia’s being an administrative city, there is nonetheless a bland anonymity built into the very structure of the capital. The population simply lacks the small accessible spaces that they could colonize and stamp with the character of their activity, as they have done historically in Rio and São Paulo. To be sure, the inhabitants of Brasilia haven’t had much time to modify the city through their practices, but the city is designed to be fairly recalcitrant to their efforts.

“Brasilité,” as a term, also underscores how the built environment affects those who dwell in it. Compared to life in Rio and São Paulo, with their color and variety, the daily round in bland, repetitive, austere Brasilia must have resembled life in a sensory deprivation tank. The recipe for high-modernist urban planning, while it may have created formal order and functional segregation, did so at the cost of a sensorily impoverished and monotonous environment—an environment that inevitably took its toll on the spirits of its residents.

The anonymity induced by Brasilia is evident from the scale and exterior of the apartments that typically make up each residential superquadra. For superquadra residents, the two most frequent complaints are the sameness of the apartment blocks and the isolation of the residences (“In Brasilia, there is only house and work”). The facade of each block is strictly geometric and egalitarian. Nothing distinguishes the exterior of one apartment from another; there are not even balconies that would allow residents to add distinctive touches and create semipublic spaces.

Brasilia is interesting only insofar as it was an entire High Modernist planned city. In most places, the Modernists rarely got their hands on entire cities at once. They did build a number of suburbs, neighborhoods, and apartment buildings. There was, however, a disconnect. Most people did not want to buy a High Modernist house or live in a High Modernist neighborhood. Most governments *did* want to fund High Modernist houses and neighborhoods, because the academics influencing them said it was the modern scientific rational thing to do. So in the end, one of High Modernists’ main contributions to the United States was the projects – ie government-funded public housing for poor people who didn’t get to choose where to live.

I never really “got” Jane Jacobs. I originally interpreted her as arguing that it was great for cities to be noisy and busy and full of crowds, and that we should build neighborhoods that are confusing and hard to get through to force people to interact with each other and prevent them from being able to have privacy, and no one should be allowed to live anywhere quiet or nice. As somebody who (thanks to the public school system, etc) has had my share of being forced to interact with people, and of being placed in situations where it is deliberately difficult to have any privacy or time to

myself, I figured Jane Jacobs was just a jerk.

But Scott has kind of made me come around. He rehabilitates her as someone who was responding to the very real excesses of High Modernism. She was the first person who really said "Hey, maybe people *like* being in cute little neighborhoods". Her complaint wasn't really against privacy or order per se as it was against extreme High Modernist perversions of those concepts that people empirically hated. And her background makes this all too understandable – she started out as a journalist covering poor African-Americans who lived in the projects and had some of the same complaints as Brazilians.

Her critique of Le Corbusierism was mostly what you would expect, but Scott extracts some points useful for their contrast with the Modernist points earlier:

First, existing structures are evolved organisms built by people trying to satisfy their social goals. They contain far more wisdom about people's needs and desires than anybody could formally enumerate. Any attempt at urban planning should try to build on this encoded knowledge, not detract from it.

Second, man does not live by bread alone. People don't want the right amount of Standardized Food Product, they want social interaction, culture, art, coziness, and a host of other things nobody will ever be able to calculate. Existing structures have already been optimized for these things, and unless you're really sure you understand all of them, you should be reluctant to disturb them.

Third, solutions are local. Americans want different things than Africans or Indians. One proof of this is that New York looks different from Lagos and from Delhi. Even if you are the world's best American city planner, you should be very concerned that you have no idea what people in Africa need, and you should be very reluctant to design an African city without extensive consultation of people who understand the local environment.

Fourth, even a very smart and well-intentioned person who is on board with points 1-3 will never be able to produce a set of rules. Most of people's knowledge is implicit, and most rule codes are quickly replaced by informal systems of things that work which are much more effective (the classic example of this is [work-to-rule strikes](#)).

Fifth, although well-educated technocrats may understand principles which give them some advantages in their domain, they are hopeless without the on-the-ground experience of the people they are trying to serve, whose years of living in their environment and dealing with it every day have given them a deep practical knowledge which is difficult to codify.

How did Jacobs herself choose where to live? As per her Wikipedia page:

[Jacobs] took an immediate liking to Manhattan's Greenwich Village, which did not conform to the city's grid structure.

V

The same thing that happened with cities happened with farms. The American version was merely farce:

We should recognize that the rationalization of farming on a huge, even national, scale was part of a faith shared by social engineers and agricultural planners throughout the world. And they were conscious of being engaged in a common endeavor... They kept in touch through journals, professional conferences, and exhibitions. The connections were strongest between American agronomists and their Russian colleagues –

connections that were not entirely broken even during the Cold War. Working in vastly different economic and political environments, the Russians tended to be envious of the level of capitalization, particularly in mechanization, of American farms while the Americans were envious of the political scope of Soviet planning. The degree to which they were working together to create a new world of large-scale, rational, industrial agriculture can be judged by this brief account of their relationship [...]

Many efforts were made to put this faith to the test. Perhaps the most audacious was the Thomas Campbell "farm" in Montana, begun – or, perhaps I should say, founded – in 1918. It was an industrial farm in more than one respect. Shares were sold by prospectuses describing the enterprise as an "industrial opportunity"; J. P. Morgan, the financier, helped to raise \$2 million from the public. The Montana Farming Corporation was a monster wheat farm of ninety-five thousand acres, much of it leased from four Native American tribes. Despite the private investment, the enterprise would never have gotten off the ground without help and subsidies from the Department of Interior and the United States Department of Agriculture (USDA).

Proclaiming that farming was about 90 percent engineering and only 10 percent agriculture, Campbell set about standardizing as much of his operation as possible. He grew wheat and flax, two hardy crops that needed little if any attention between planting and harvest time. The land he farmed was the agricultural equivalent of the bulldozed site of Brasilia. It was virgin soil, with a natural fertility that would eliminate the need for fertilizer. The topography also vastly simplified matters: it was flat, with no forests, creeks, rocks, or ridges that would impede the smooth course of machinery over its surface. In other words, the selection of the simplest, most standardized crops and the leasing of something very close to a blank agricultural space were calculated to favor the application of industrial methods [...]

This is not the place to chronicle the fortunes of the Montana Farming Corporation, and in any event Deborah Fitzgerald has done so splendidly. Suffice it to note that a drought in the second year and the elimination of a government support for prices the following year led to a collapse that cost J. P. Morgan \$1 million. The Campbell farm faced other problems besides weather and prices: soil differences, labor turnover, the difficulty of finding skilled, resourceful workers who would need little supervision. Although the corporation struggled on until Campbell's death in 1966, it provided no evidence that industrial farms were superior to family farms in efficiency and profitability.

But the Soviet version was tragedy. Instead of raising some money to start a giant farm and seeing it didn't work, the USSR uprooted millions of peasants, forced them onto collective farms, and then watched as millions of people starved to death due to crop failure. What happened?

Scott really focuses on that claim (above) that farming was "90% engineering and only 10% agriculture". He says that these huge farms all failed – despite being better-funded, higher-tech, and having access to the wisdom of the top agricultural scientists – exactly because this claim was false. Small farmers may not know much about agricultural science, but they know a lot about farming. Their knowledge is intuitive and local – for example, what to do in a particular climate or soil. It is sometimes passed down over generations, and other times determined through long hours of trial-and-error.

He gave the example of Tanzania, where small farmers grew dozens of different crops together in seeming chaos. Western colonists tried to convince them – often by force – to switch to just growing one thing at a time to reap advantages of efficiency, standardization, and specialization of labor. Only growing one crop in the same field was Agricultural Science 101. But this turned out to be a bad idea in the difficult Tanzanian environment:

The multistoried effect of polyculture has some distinct advantages for yields and soil conservation. "Upper-story" crops shade "lowerstory" crops, which are selected for their ability to thrive in the cooler soil

temperature and increased humidity at ground level. Rainfall reaches the ground not directly but as a fine spray that is absorbed with less damage to soil structure and less erosion. The taller crops often serve as a useful windbreak for the lower crops. Finally, in mixed or relay cropping, a crop is in the field at all times, holding the soil together and reducing the leaching effects that sun, wind, and rain exert, particularly on fragile land. Even if polyculture is not to be preferred on the grounds of immediate yield, there is much to recommend it in terms of sustainability and thus long-term production.

Our discussion of mixed cropping has thus far dealt only with the narrow issues of yield and soil conservation. It has overlooked the cultivators themselves and the various other ends that they seek by using such techniques. The most significant advantage of intercropping, Paul Richards claims, is its great flexibility, “the scope [it] offers for a range of combinations to match individual needs and preferences, local conditions, and changing circumstances within each season and from season to season.” Farmers may polycrop in order to avoid labor bottlenecks at planting and at harvest.⁴⁴ Growing many different crops is also an obvious way to spread risks and improve food security. Cultivators can reduce the danger of going hungry if they sow, instead of only one or two cultivars, crops of long and short maturity, crops that are drought resistant and those that do well under wetter conditions, crops with different patterns of resistance to pests and diseases, crops that can be stored in the ground with little loss (such as cassava), and crops that mature in the “hungry time” before other crops are gathered. Finally, and perhaps most important, each of these crops is embedded in a distinctive set of social relations. Different members of the household are likely to have different rights and responsibilities with respect to each crop. The planting regimen, in other words, is a reflection of social relations, ritual needs, and culinary tastes; it is not just a production strategy that a profit-maximizing entrepreneur took straight out of the pages of a text in neoclassical economics.

Nor could this be solved just by adding a pinch of empiricism. A lot of European farming specialists were into empiricism, sort of. What they ended up doing was creating crops that worked really well in a lab but not in actual Tanzania. If they were lucky, they created crops that worked really well on the one experimental farm in Tanzania they fenced off as a testing ground, but not on any other Tanzanian farms. If they were *really* lucky, they created crops that would grow on Tanzanian farms and be good on whatever single axis they were optimizing (like selling for lots of money) but not in other ways that were equally important to the populace (like being low-risk, or useful for non-food purposes, or resistant to disease, or whatever). And if they were *supremely* lucky, then they would go to the Tanzanians and say “Hey, we invented a new farming method that solves all your problems!” and the Tanzanians would say “Yeah, we heard rumors about that, so we tried it ourselves, and now we’ve been using it for five years and advanced it way beyond what you were doing.”

There *were* some scientists who got beyond these failure modes, and Scott celebrates them (while all too often describing how they were marginalized and ignored by the rest of the scientific community). But at the point where you’ve transcended all this, you’re no longer a domain-general agricultural scientist, you’re a Tanzanian farming specialist who’s only one white coat removed from being a Tanzanian farmer yourself.

Even in less exotic locales like Russia, the peasant farmers were extraordinary experts on the conditions of their own farms, their own climates, and their own crops. Take all of these people, transport them a thousand miles away, and give them a perfectly rectangular grid to grow Wheat Cultivar #6 on, and you have a recipe for disaster.

VI

So if this was such a bad idea, why did everyone keep doing it?

Start with the cities. Scott notes that although citizens generally didn’t have a problem with earlier cities, governments

did:

Historically, the relative illegibility to outsiders of some urban neighborhoods has provided a vital margin of political safety from control by outside elites. A simple way of determining whether this margin exists is to ask if an outsider would have needed a local guide in order to find her way successfully. If the answer is yes, then the community or terrain in question enjoys at least a small measure of insulation from outside intrusion. Coupled with patterns of local solidarity, this insulation has proven politically valuable in such disparate contexts as eighteenth-and early nineteenth-century urban riots over bread prices in Europe, the Front de Liberation Nationale's tenacious resistance to the French in the Casbah of Algiers, and the politics of the bazaar that helped to bring down the Shah of Iran. Illegibility, then, has been and remains a reliable resource for political autonomy

This was a particular problem in Paris, which was famous for a series of urban insurrections in the 19th century (think *Les Misérables*, but about once every ten years or so). Although these generally failed, they were hard to suppress because locals knew the “terrain” and the streets were narrow enough to barricade. Slums full of poor people gathered together formed tight communities where revolutionary ideas could easily spread. The late 19th-century redesign of Paris had the explicit design of destroying these areas and splitting up poor people somewhere far away from the city center where they couldn't do any harm.

The Soviet collective farms had the same dubious advantage. The problem they “effectively” “solved” was the non-collectivized farmers becoming too powerful and independent a political bloc. They lived in tight-knit little villages that did their own thing, the Party officials who went to these villages to keep order often ended up “going native”, and the Soviets had no way of knowing how much food the farmers were producing and whether they were giving enough of it to the Motherland:

Confronting a tumultuous, footloose, and “headless” rural society which was hard to control and which had few political assets, the Bolsheviks, like the scientific foresters, set about redesigning their environment with a few simple goals in mind. They created, in place of what they had inherited, a new landscape of large, hierarchical, state-managed farms whose cropping patterns and procurement quotas were centrally mandated and whose population was, by law, immobile. The system thus devised served for nearly sixty years as a mechanism for procurement and control at a massive cost in stagnation, waste, demoralization, and ecological failure.

The collectivized farms couldn't grow much, but people were thrown together in artificial towns designed to make it impossible to build any kind of community: there was nowhere to be except in bed asleep, working in the fields, or at the public school receiving your daily dose of state propaganda. The towns were identical concrete buildings on a grid, which left the locals maximally disoriented (because there are no learnable visual cues) and the officials maximally oriented (because even a foreigner could go to the intersection of Street D and Street 7). All fields were perfectly rectangular and produced Standardized Food Product, so it was (theoretically) easy to calculate how much they should be producing and whether people were meeting that target. And everyone was in the same place, so if there were some sort of problem it was much easier to bring in the army or secret police than if they were split up among a million tiny villages in the middle of nowhere.

So although modernist cities and farms may have started out as attempts to help citizens with living and farming, they ended up as contributors to the great government project of legibility and taxing people effectively. *Seeing Like A State* summarizes the sort of on-the-ground ultra-empirical knowledge that citizens have of city design and peasants of farming as *metis*, a Greek term meaning “practical wisdom”. I was a little concerned about this because they seem like two different things. The average citizen knows nothing about city design and in fact does not design cities; cities sort of happen in a weird way through cultural evolution or whatever. The average farmer knows a lot about farming (even

if it is implicit and not as book learning) and applies that knowledge directly in how they farm. But Scott thinks these are more or less the same thing, that this thing is a foundation of successful communities and industries, and that ignoring and suppressing it is what makes collective farms and modernist planned cities so crappy. He generalizes this further to almost every aspect of a society – its language, laws, social norms, and economy. But this is all done very quickly, and I feel like there was a sleight of hand between “each farmer eventually figures out how to farm well” and “social norms converge on good values”.

Insofar as Scott squares the above circle, he seems to think that many actors competing with each other will eventually carve out a beneficial equilibrium better than that of any centralized authority. This doesn’t really mesh will with my own fear that many actors competing with each other will eventually shoot themselves in the foot and destroy everything, and I haven’t really seen a careful investigation of when we get one versus the other.

VII

What are we to make of all of this?

Well, for one thing, Scott basically admits to stacking the dice against High Modernism and legibility. He admits that the organic livable cities of old had life expectancies in the forties because nobody got any light or fresh air and they were all packed together with no sewers and so everyone just died of cholera. He admits that at some point agricultural productivity multiplied by like a thousand times and the Green Revolution saved millions of lives and all that, and probably that has something to do with scientific farming methods and rectangular grids. He admits that it’s pretty convenient having a unit of measurement that local lords can’t change whenever they feel like it. Even modern timber farms seem pretty successful. After all those admissions, it’s kind of hard to see what’s left of his case.

(also, I grew up in Irvine, the most planned of planned cities, and I loved it.)

What Scott eventually says is that he’s not against legibility and modernism per se, but he wants to present them as ingredients in a cocktail of state failure. You need a combination of four things to get a disaster like Soviet collective farming (or his other favorite example, compulsory village settlement in Tanzania). First, a government incentivized to seek greater legibility for its population and territory. Second, a High Modernist ideology. Third, authoritarianism. And fourth, a “prostrate civil society”, like in Russia after the Revolution, or in colonies after the Europeans took over.

I think his theory is that the back-and-forth between centralized government and civil society allows scientific advances to be implemented smoothly instead of just plowing over everyone in a way that leads to disaster. I also think that maybe a big part of it is incremental versus sudden: western farming did well because it got to incrementally add advances and see how they worked, but when you threw the entire edifice at Tanzania it crashed and burned.

I’m still not really sure what’s left. Authoritarianism is bad? Destroying civil society is bad? You shouldn’t do things when you have no idea what you’re doing and all you’ve got to go on is your rectangle fetish? The book contained some great historical tidbits, but I’m not sure what overarching lesson I learned from it.

It’s not that I don’t think Scott’s preference for *metis* over scientific omnipotence has value. I think it has lots of value. I see this all the time in psychiatry, which always has been and to some degree still is really High Modernist. We are educated people who know a lot about mental health, dealing with a poor population who (in the case of one of my patients) refers to Haldol as “Hound Dog”. It’s very easy to get in the trap of thinking that you know better than these people, especially since you often do (I will never understand how many people are *shocked* when I diagnose their sleep disorder as having something to do with them drinking fifteen cups of coffee a day).

But psychiatric patients have a *metis* of dealing with their individual diseases the same way peasants have a *metis* of dealing with their individual plots of land. My favorite example of this is doctors who learn their patients are taking marijuana, refuse to keep prescribing them their vitally important drugs unless the patient promises to stop, and then gets surprised when the patients end up decompensating because the marijuana was keeping them together. I'm not saying smoking marijuana is a good thing. I'm saying that for some people it's a load-bearing piece of their mental edifice. And if you take it away without any replacement they will fall apart. And *they have explained this to you a thousand times and you didn't believe them.*

There are *so many* fricking patients who respond to sedative medications by becoming stimulated, or stimulant medications by becoming sedated, or who become more anxious whenever they do anti-anxiety exercises, or who hallucinate when placed on some super common medication that has never caused hallucinations in anyone else, or who become suicidal if you try to reassure them that things aren't so bad, or any other completely perverse and ridiculous violation of the natural order that you can think of. And the only redeeming feature of all of this is that the patients themselves know all of this stuff super-well and are usually happy to tell you if you ask.

I can totally imagine going into a psychiatric clinic armed with the Evidence-Based Guidelines the same way Le Corbusier went into Moscow and Paris armed with his Single Rational City Plan and the same way the agricultural scientists went into Tanzania armed with their List Of Things That Definitely Work In Europe. I expect it would have about the same effect for about the same reason.

(including [the part where I would get promoted](#). I'm not too sure what's going on there, actually.)

So fine, Scott is completely right here. But I'm only bringing this up because it's something I've already thought about. If I didn't already believe this, I'd be indifferent between applying the narrative of the wise Tanzanian farmers knowing more than their English colonizers, versus the narrative of the dumb yokels who refuse to get vaccines because they might cause autism. [Heuristics work until they don't](#). Scott provides us with these great historical examples of local knowledge outdoing scientific acumen, but other stories present us with great historical examples of the opposite, and when to apply which heuristic seems really unclear. Even "don't bulldoze civil society and try to change everything at once" goes astray sometimes; the Meiji Restoration was wildly successful by doing exactly that.

Maybe I'm trying to take this too far by talking about psychiatry and Meiji Restorations. Most of Scott's good examples involved either agriculture or resettling peasant villages. This is understandable; Scott is a scholar of colonialism in Southeast Asia and there was a lot of agriculture and peasant resettling going on there. But it's a pretty limited domain. The book amply proves that peasants know an astounding amount about how to deal with local microclimates and grow local varieties of crops and so on, and frankly I am shocked that anyone with an IQ of less than 180 has ever managed to be a peasant farmer, but how does that apply to the sorts of non-agricultural issues we think about more often?

The closest analogy I can think of right now – maybe because it's on my mind – is [this story about check-cashing shops](#). Professors of social science think these shops are evil because they charge the poor higher rates, so they should be regulated away so that poor people don't foolishly shoot themselves in the foot by going to them. But on closer inspection, they offer a better deal for the poor than banks do, for complicated reasons that aren't visible just by comparing the raw numbers. Poor people's understanding of this seems a lot like the *metis* that helps them understand local agriculture. And progressives' desire to shift control to the big banks seems a lot like the High Modernists' desire to shift everything to a few big farms. Maybe this is a point in favor of something like libertarianism? Maybe especially a "libertarianism of the poor" focusing on things like occupational licensing, not shutting down various services to the poor because they don't meet rich-people standards, not shutting down various services to the poor because we think they're "price-gouging", et cetera?

Maybe instead of concluding that Scott is too focused on peasant villages, we should conclude that he's focused on

confrontations between a well-educated authoritarian overclass and a totally separate poor underclass. Most modern political issues don't exactly map on to that – even things like taxes where the rich and the poor are on separate sides don't have a bimodal distribution. But in cases there are literally about rich people trying to dictate to the poorest of the poor how they should live their lives, maybe this becomes more useful.

Actually, one of the best things the book did to me was make me take cliches about "rich people need to defer to the poor on poverty-related policy ideas" more seriously. This has become so overused that I roll my eyes at it: "Could quantitative easing help end wage stagnation? Instead of asking macroeconomists, let's ask this 19-year old single mother in the Bronx!" But Scott provides a lot of situations where that was exactly the sort of person they should have asked. He also points out that Tanzanian natives using their traditional farming practices were more productive than European colonists using scientific farming. I've had to listen to so many people talk about how "we must respect native people's different ways of knowing" and "native agriculturalists have a profound respect for the earth that goes beyond logocentric Western ideals" and nobody had ever bothered to tell me before that they *actually produced more crops per acre*, at least some of the time. That would have put all of the other stuff in a pretty different light.

I understand Scott is an anarchist. He didn't really try to defend anarchism in this book. But I was struck by his description of peasant villages as this totally separate unit of government which was happily doing its own thing very effectively for millennia, with the central government's relevance being entirely negative – mostly demanding taxes or starting wars. They kind of reminded me of some pictures of hunter-gatherer tribes, in terms of being self-sufficient, informal, and just never encountering the sorts of economic and political problems that we take for granted. They make communism (the type with actual communes, not the type where you have Five Year Plans and Politburos and gulags) look more attractive. I think Scott was trying to imply that this is the sort of thing we could have if not for governments demanding legibility and a world of universal formal rule codes accessible from the center? Since he never actually made the argument, it's hard for me to critique it. And I wish there had been more about cultural evolution as separate from the more individual idea of *metis*.

A final note: Scott often used the word "rationalism" to refer to the excesses of High Modernism, and I've deliberately kept it. What relevance does this have for the LW-Yudkowsky-Bayesian rationalist project? I think the similarities are more than semantic; there certainly is a hope that learning domain-general skills will allow people to leverage raw intelligence and The Power Of Science to various different object-level domains. I continue to be doubtful that this will work in the sort of practical domains where people have spent centuries gathering *metis* in the way Scott describes; this is why I'm wary of any attempt of the rationality movement to branch into self-help. I'm more optimistic about rationalists' ability to open underexplored areas like existential risk – it's not like there's a population of Tanzanian peasants who have spent the last few centuries developing traditional x-risk research whom we are arrogantly trying to replace – and to focus on things that don't bring any immediate practical gain but which help build the foundations for new philosophies, better communities, and more positive futures. I also think that a good art of rationality would look a lot like *metis*, combining easily teachable mathematical rules with more implicit virtues which get absorbed by osmosis.

Overall I did like this book. I'm not really sure what I got from its thesis, but maybe that was appropriate. *Seeing Like A State* was arranged kind of like the premodern forests and villages it describes; not especially well-organized, not really directed toward any clear predetermined goal, but full of interesting things and lovely to spend some time in.

Book Review: Legal Systems Very Different From Ours

Posted on November 13, 2017 by Scott Alexander



I

Medieval Icelandic crime victims would sell the right to pursue a perpetrator to the highest bidder. 18th century English justice replaced fines with criminals bribing prosecutors to drop cases. Somali judges compete on the free market; those who give bad verdicts get a reputation that drives away future customers.

“Anarcho-capitalism” evokes a dystopian cyberpunk future. But maybe that’s wrong. Maybe we’ve always been anarcho-capitalist. Maybe a state-run legal system isn’t a fact of nature, but a historical oddity as contingent as collectivized farming or nationalized railroads. [Legal Systems Very Different From Ours](#), by anarcho-capitalist/legal scholar/medieval history buff David Friedman, successfully combines the author’s three special interests into a whirlwind tour of exotic law.

Law is a public good. Crime victims have little economic incentive to punish the perpetrator; if you burn my house down, jailing you won’t un-burn the house. If you steal my gold, I have some interest in catching you and taking it back, but no more than I do in catching some other poor shmuck and taking *his* gold. It’s only society as a whole that wants to make sure criminals are reliably punished and the innocent consistently safe. This is the classic situation where economists usually recommend government intervention.

But sometimes that doesn’t work. Maybe you live in an area like Somalia or medieval Ireland without a strong centralized government. Maybe you live in a strato-klepto-kakocracy run by warlords who can’t even pronounce “jurisprudence”, let alone enforce it. Maybe you’re a despised minority group whom the State wants nothing to do with, or who wants nothing to do with the State.

Gypsies living scattered in foreign countries have generally wanted to run their own communities by their own rules. Nothing stops some of them from calling themselves a “legislature” or a “court” and claiming to make laws or pass sentences. But something does stop them from trying to enforce them: from the State’s point of view, a “court” that executes an offender is just a bunch of Gypsies who got together and committed murder. So the Vlach Rom – Romanian Gypsies – organize courts called *kris* which enforce their sentences with threat of banishment from the community.

Gypsies traditionally believe in *marime*, a sort of awful pollution that infects people who don’t follow the right rituals; anyone who interacts with polluted people will become polluted themselves. *Kris* courts can declare the worst offenders polluted, ensuring their speedy ostracization from Gypsy society. And since non-Gypsies are polluted by default, the possibility of ostracism and forced integration into non-Gypsy society will seem intolerable:

The effectiveness of that threat [of ostracism] depends on how easily the exiled gypsy can function outside of his community. The marimé rules (and similar rules in other societies) provide a mechanism for isolating the members of the community. Gaije, non-gypsies, do not know the marimé rules and so do not and cannot obey them. It follows that they are all polluted, unclean, carriers of a contagious disease, people whom no Rom in his right mind would willingly choose to associate with; when and if such association is unavoidable it must be taken with great care. The gypsy view of gaije, reinforced by the gaije view of gypsies as

uneducated and illiterate thieves and swindlers, eliminates the exit option and so empowers the kris to enforce gypsy law by the threat of exclusion from the only tolerable human society.

This reminds me of [The Use And Abuse Of Witchdoctors For Life](#): once your culture has a weird superstition, it can get plugged into various social needs to become a load-bearing part of the community structure.

Amish also live under the authority of a foreign culture and have settled on a similar system, with a twist. The basic unit of Amish society is the church congregation; Amish settlements big enough to support multiple churches will have many congregations mixed together. Each congregation will have its own rules, especially about which technologies their members are or aren't allowed to use. Amish people who violate their congregation's rules, either by using forbidden technology or by the usual litany of sins, are punished with public confession or temporary ostracism. Amish people who refuse to abide by lesser punishments are excommunicated, though they can be un-excommunicated if they change their minds and agree to follow the court's orders.

Amish congregations are nominally democratic, but in practice Friedman calls them dictatorship-like because everyone votes the way the bishop wants. But they are a "competitive dictatorship"; since there are so many different congregations in the same town, an Amish family who doesn't like their congregation's leadership or legal system can move to another congregation and agree to be bound by their laws instead. This makes it a rare remaining example of a polycentric legal system outside anarcho-capitalist fantasies or [Too Like The Lightning](#):

Such a system can be viewed as a competitive market for legal rules, constrained, like other competitive markets, to produce about the product that the customers want. Competitive dictatorship is the mechanism we routinely use to control hotels and restaurants; the customers have no vote on what color the walls are painted or what is on the menu, but an absolute vote on which one they patronize.

They do encounter the same problem as the Gypsies: can you just commit a crime, then accept your ostracism and integrate with another society somewhere else? The Amish have some internal mechanisms to prevent this: congregations are usually on good terms with each other, but if Congregation A accepts a member being shunned by Congregation B, then all of Congregation B's members will shun all of Congregation A's members. In practice, this makes it easy to switch rules as a member in good standing who honestly doesn't like the laws, but hard to break the laws and get away with it.

Of course, you can still leave the Amish community and go join broader American society. But have you *seen* broader American society?

18th century England had a government, a court system, and some minimal law enforcement – but it really sucked. There were no public prosecutors; anyone who felt like it could bring a criminal to court and start prosecuting him, but if nobody felt like it then the crime remained unpunished. Prosecuting took a lot of time and money and was generally a thankless task. And the government didn't want to go to the expense of imprisoning people, so they usually just hanged convicted offenders (if the crime seemed really bad) or pardoned them (if it didn't seem to merit hanging). The exotic anarcho-capitalist part comes in as English civil society creates its own structures to work around these limitations.

Merchants, landowners, and other people with wealth banded together in mutual-protection-insurance-groups. Everyone in the group would pay a fixed amount yearly, and if one of them got robbed the group would use the money to hire a prosecutor to try the criminal. Group members would publish their names in the newspaper to help inform thieves whom it was a bad idea to rob. But this wasn't about leaving poor people out to dry. The groups would also help indigents who couldn't afford their own prosecutors, partly out of a desire to crack down on crime before it reached the point where it could inconvenience them. They wouldn't help people who could have afforded insurance

but declined anyway, though – otherwise there would be no incentive to buy in.

(if this sounds familiar, it's from [another, very different David Friedman book](#))

What about the lack of good punishments? Once a trial was underway, prosecutors would usually cut a deal: the offender would bribe the prosecutor with a certain amount, and the prosecutor would drop the case. The size of the bribe would vary based on how much the offender could pay, the extent of their crime, and the facts of the case (and therefore the likelihood of the magistrate choosing hanging vs. pardon). This not only helped tailor the punishment more precisely to the crime, but helped defer the cost of prosecution: victims (or their mutual-protection-insurance-groups) were incentivized to press charges because they could recoup their costs through the bribes paid to drop them:

What both modern and contemporary commentators seem to have missed is that, however corrupt such arrangements might be from a legal standpoint, they helped solve the fundamental problem of private prosecution. The possibility of compounding provided an incentive to prosecute—it converted the system into something more like a civil system, where a victim sues in the hope of collecting money damages. And while compounding might save the criminal from the noose, he did not get off scott-free. He ended up paying, to the prosecutor, what was in effect a fine.

10th through 13th century Iceland was in the same position as the Vlach Rom: a legislature (the Althing), some courts, but no executive branch. Unlike the Rom, the Icelanders' problem wasn't foreign oppressors – it was that they were the Viking equivalent of those hard-core libertarians who live in compounds in Montana where the Feds can't reach them. In this case "the Feds" were the forces of King Harald Fairhair, who had just taken over and centralized power in Norway. Some Norwegians decided they would rather live on a remote and frequently-exploding piece of rock on the edge of the world than be anyone's subject: thus, medieval Iceland.

If an Icelander thought a crime had happened, they would go to court and plead the case themselves. If the court pronounced a guilty verdict, it would demand a penalty from the criminal. Usually this was a fine paid to the victim; even murders were punished with wergeld. If the criminal paid the fine voluntarily, all was well. If they refused – or didn't even come to court – then the court could declare the criminal an outlaw, meaning it was legal to kill him and take his stuff. And:

One obvious objection to a system of private enforcement is that the poor (or weak) would be defenseless. The Icelandic system dealt with this problem by giving the victim a property right – the right to be reimbursed by the criminal – and making that right transferable. The victim could turn over his case to someone else, either gratis or in return for a consideration. A man who did not have sufficient resources to prosecute a case or enforce a verdict could sell it to another who did and who expected to make a profit in both money and reputation by winning the case and collecting the fine. This meant that an attack on even the poorest victim could lead to eventual punishment.

A second objection is that the rich (or powerful) could commit crimes with impunity, since nobody would be able to enforce judgment against them. Where power is sufficiently concentrated this might be true; this was one of the problems which led to the eventual breakdown of the Icelandic legal system in the thirteenth century. But so long as power was reasonably dispersed, as it seems to have been for the first two centuries after the system was established, this was a less serious problem. A man who refused to pay his fines was outlawed and would probably not be supported by as many of his friends as the plaintiff seeking to enforce judgment, since in case of violent conflict his defenders would find themselves legally in the wrong. If the lawbreaker defended himself by force, every injury inflicted on the partisans of the other side would result in another suit, and every refusal to pay another fine would pull more people into the coalition against him.

There is a scene in Njal's Saga that provides striking evidence of the stability of this system. Conflict between two groups has become so intense that open fighting threatens to break out in the middle of the court. A leader of one faction asks a benevolent neutral what he will do for them in case of a fight. He replies that if they are losing he will help them, and if they are winning he will break up the fight before they kill more men than they can afford! Even when the system seems so near to breaking down, it is still assumed that every enemy killed must eventually be paid for. The reason is obvious enough; each man killed will have friends and relations who are still neutral—and will remain neutral if and only if the killing is made up for by an appropriate wergeld.

I think this is asking: are we sure you can't end up with outlaw cascades, where everyone just agrees to be outlaws together? Suppose Warren Buffett cuts off my arm. The court asks him to pay a fine, and he refuses, so the court declares him an outlaw and legally killable. I gather some of my friends to form a posse to kill him, but he hires a hundred bodyguards to resist me. There's a fight, the bodyguards kill my friends, and the court fines the bodyguards. They don't pay, so the court declares the bodyguards outlaws. I gather a thousand people to kill Buffett and/or his hundred bodyguards, and Buffett and his bodyguards pool their money to hire a whole force of mercenaries to resist us. The mercenaries kill lots of us, the court fines them, and the mercenaries don't pay. Now the court declares the mercenaries outlaws. But it seems like at some point maybe more than half the population of Iceland will be outlaws, and then maybe they just have to declare a new legal system or something.

An Icelander might retort: why doesn't that happen in modern America? A policeman catches you dealing drugs, so you offer the policeman \$10,000 to let it pass. The policeman refuses because it's illegal and he would get in trouble. Well, you say, what's the worst thing that could happen if you got in trouble? The police would come after you? But police would hesitate to arrest a fellow officer, plus we've already established that they can be deflected with bribes. Sure, there's a stable equilibrium where you arrest me right now. But there's *also* a stable equilibrium where 51%+ of the nation's police join our sordid bribery chain, accumulate more power than the law-abiding police, and end up as some weird mercenary army that takes over the country and rewrites the law to their own advantage.

This is a good place to remember that David Friedman is also the author of [A Positive Account Of Property Rights](#), maybe the single most mind-opening essay I've ever read. No summary can do it justice, but the basic outline is that governmental "legitimacy" is the government's position as a conspicuous Schelling point for everybody who wants to avoid civil war/the state of nature/a worse government. Once it's common knowledge that a government is legitimate, everyone expects everyone else to enforce its rules, and so they'll enforce its rules in turn until it becomes common knowledge that the government isn't legitimate anymore. This works just as well in medieval Icelandic anarcho-capitalism as it does in modern America. Just because our government dresses all of its enforcers-of-state-sanctioned violence in snazzy uniforms and makes them work out of the same building doesn't make the whole system any less of a mass hallucination.

II

This book works well alongside James Scott's [Seeing Like A State](#) and the whole discourse around [cultural evolution](#).

In *Seeing Like A State*, ordinary people living their daily lives blunder into highly advanced systems for doing whatever it is they do. Primitive farmers will know every tiny detail about exactly when to plant which crops, and how to exploit microvariations in soil quality, and know ridiculous tricks like planting fish heads in the ground as fertilizer. Ordinary city-dwellers will organically build houses and stores and streets in exactly the right fractal patterns to maximize some measure of quality of life. Scott dubs this "metis", an evolved intuitive sense of practical wisdom that often outperforms seemingly more scientific solutions.

Many of the societies Friedman profiles in *Legal Systems Very Different From Ours* seem to operate on *metis*. Most don't

know who developed their legal system; in a few of them, it is explicitly declared to have been the work of God. Most don't really know why their legal system works – in some cases, Friedman only gives an economic analysis of why some rule might exist after admitting that previous scholarship (both modern academic, and within the society in question) has failed to come up with answers. And a lot of them are too brilliant, and need too many weird interlocking parts, to be the work of any single person.

"Cultural evolution" is the idea that cultures evolve in a way analogous to biological organisms. The definition gets kind of fuzzy – if I come up with a good idea and my culture adopts it, is that the result of "cultural evolution" or ordinary human ingenuity? But a lot of people find the concept to have some value – and if it has any at all, *Legal Systems Very Different From Ours* has to include some of the best examples.

Friedman frames this in economic terms. Social "entrepreneurs" come up with some new system that solves a need, and it catches on by raising the utility of everyone involved. The mutual-protection-insurance-groups of 18th century England work this way: somebody invents them and offers the opportunity for other people to sign on, everyone who does ends up better off than the people who doesn't, and they eventually reach fixation. Same with the criminal-prosecutor bribes; someone thinks it up, it leaves both sides better off, so everybody who hears about it does it. Viewed very optimistically, wherever there's a problem in your culture, institutions to solve the problem will magically appear and spread until everybody does them.

Conflict is an especially fertile ground for cultural innovation. Friedman stresses how many legal systems, including advanced ones with lawyers and codes and everything, show signs of originating from feud systems, which might be the most basic form of law. They work like this: "If you offend me in some way, I will try to kill you". A slightly more advanced version that takes account of possibly power differentials between offender and victim: "If you offend me in some way, everybody in my family will try to kill everybody in your family". This originally sounds unpromising, but it turns out that people really don't want their family members murdered. So we end up with an even more advanced version: "If you offend me in some way, we had better find some way to arbitrate our dispute, *or else* everybody in my family will try to kill everybody in your family".

The Somali system seems to be somewhere around here: if two people have a dispute, they find a mutually agreeable judge to arbitrate; the judge will decide who's in the wrong and what fine they need to pay to make it right. If someone refuses to go to the judge, or refuses to abide by the judge's decision, then it's family-member-killing time. Needless to say, Somali judges' services remain popular. And since judges gain status by arbitrating, and since only judges who make widely-regarded-as-good decisions get invited to keep doing so, there's economic pressure for the judges to make good decisions (which then go down as precedent and inspire future cases). It's easy to see how something like this can turn into a perfectly respectable legal system where people totally forget that killing each other's family members is even an option. Catch it at this last stage, and hear enough people admit they have no idea who "invented" their legal system, and it looks like it appeared by magic.

In fact, one of the most interesting things I got from this book is that all legal systems need a punishment of last resort – one that can be enforced whether or not the offender agrees with it – but these punishments practically never happen in real life. The Gypsies and Amish will ostracize members who defy the court – but since everyone lives in fear of ostracization, in real life they'll just pay the fine or make their public confession or whatever. The English will hang criminals at the drop of a hat – but since the threat of hanging incentivizes them to bribe prosecutors, in reality few people will need to be hanged. The Icelandic courts could declare offenders outlaws who can be killed without repercussion – but the threat encourages Icelanders to pay the *vergeld*, and nobody has to get outlawed. The Somalis are ready to have murderous family feuds – but the possibility of such a feud keeps people willing to go to arbitration. Even our own legal system works like this. The police can physically drag you to jail, kicking and screaming. But more likely you're going to plea bargain, or agree to community service, or at least be cooperative and polite while the police take you away. Plea bargains – which are easier for prosecutors, easier for defendants, and easier for taxpayers – seem like a good example of cultural evolution in action; once someone thought them up, there was no way they weren't going to take over everything despite their very serious costs.

III

Three other things worth noting about *Legal Systems Very Different From Ours*.

First, something kept seeming off about all the legal systems mentioned, which only clicked into place about halfway through: they really, really didn't seem prepared for crime. A lot of them worked on a principle like: "If there's a crime, we'll call together a court made of all the town elders, plus at least three different religious leaders, plus the heads of the families of everybody involved, plus a representative of the Great King, plus nine different jurists from nine different universities, and all of them will meet on the Field Of Meeting, and a great tent will be erected, and..." The whole thing sounded like it might work as long as there was like one crime a year. Any more than that and none of the society's officials would ever have time for anything else.

As weird as it is to punish murder with a fine, the fines these societies levied for murder sounded really high: the Islamic price was a hundred camels, the Irish price was seven female slaves. The average person wouldn't have that many slaves or camels, so people in Arabia or Ireland would band together into clan/family-based blood-money-paying-groups that acted kind of like insurance companies. If a member got convicted of a crime, everyone else would come together to help them pony up the money. I assume this helped incentivize people's families to discourage them from committing crimes. But it has the same feeling of nobody expecting very many crimes to be committed. How much of medieval Arabia's GDP consisted of transfers of 100 camels from murderers to victims' families?

One little-admitted but much-worried-about justification for mass incarceration in our society is the concern that some people are just so naturally violent that, left in the outside world, they would offend again and again until they died. The societies in this book didn't seem to worry about this. If someone killed, their family would give up the relevant number of camels, and then everyone would be on their way. As far as I can tell, the Amish have *no idea* what to do about any crime more dire than using a telephone. Nobody used anything at all like incarceration. 18th century England occasionally sent prisoners somewhere horrible like America, but once the colonies revolted they experimented with jails, found them too expensive, and just sort of flailed around punishment-less until they finally discovered Australia.

There's a lot of concern about police brutality, police racism, police failure-to-actually-control crime, et cetera. A few far-leftists have flirted with the idea of abolishing police, and the only way I can make sense of this is by analogy to something like Somali or Icelandic law. These were genuine community-based non-hierarchical legal systems. And, for the place and the time, they seem to have worked really well (Somaliland, which uses traditional Somali law, is doing way better than Somalia proper, whose law system is somewhat westernized). But I also know that it's weirdly hard to get a good picture of how modern crime rates compare to ancient ones. On the one hand are statistics like the ones saying crime has increased by an order of magnitude since 1900 or so; on the other are findings like Steven Pinker's that violence is constantly declining. Apply the "court made of town elders plus at least three different religious leaders plus..." to Baltimore, and the Field Of Meeting is going to get pretty crowded. On the other hand, in my past work with criminals I've been constantly surprised by how much role their families and their communities still play in their lives, and maybe a system that left legal enforcement up to them would do better than the overstretched and underperforming police.

(but what would the transfer process look like? Just cancel all funding for the Baltimore Police Department and hope for the best?)

Second, some complaints that are kind of unfair because they're along the lines of "this book is too good", but which probably need a mention.

Whenever I read a book by anyone other than David Friedman about a foreign culture, it sounds like "The X'wunda give their mother-in-law three cows every monsoon season, then pluck out their own eyes as a sacrifice to Humunga,

the Volcano God".

And whenever I read David Friedman, it sounds like "The X'wunda ensure positive-sum intergenerational trade by a market system in which everyone pays the efficient price for continued economic relationships with their spouse's clan; they demonstrate their honesty with a costly signal of self-mutilation that creates common knowledge of belief in a faith whose priests are able to arbitrate financial disputes."

This is great, and it's important to fight the temptation to think of foreign cultures as completely ridiculous idiots who do stuff for no reason. But it all works out so neatly – and so much better than when anyone else treats the same topics – that I'm always nervous if I'm not familiar enough with the culture involved to know whether they're being shoehorned into a mold that's more rational-self-interest-maximizing than other anthropologists (or they themselves) would recognize.

And also, the cultural evolution idea is really optimistic. I've been trying to read a bit more about Marxism and Postmodernism lately, and they would be pretty skeptical about analyzing social systems by asking "What large-scale problem of human interaction is this system the exactly optimal solution for?"

Like, there's a perspective where lots of countries have a King, because societies that have a single central nexus to their coordination structure are able to coordinate better than ones that don't, and having them rule for life promotes long-term thinking, and them be hereditary provides a clear Schelling Point for secession disputes that prevents civil war *and* cleverly ensures that the previous ruler is incentivized to promote the peaceful transfer of power to the next one, and this is why [constitutional monarchies have slightly higher yearly GDP growth](#) than other forms of government.

And there's another perspective where lots of countries have a King, because some guy seized so much power that he can live in a giant palace and order people around all day instead of doing work. And if anyone tries to prevent him from doing that, he can arrange to have that person beheaded. *Legal Systems Very Different From Ours* is very much part of the first perspective. It's a story of nations and legal systems evolving towards ever-more-optimal and ever-more-efficient institutions for the good of all, and it presents strong evidence supporting that story. I can't disagree with its evidence from within its narrative, but I still wonder how much to worry about this alternate way of looking at things.

Third, in all of the fretting about how terrible our government is, and trying to change our government to be less terrible, and trying to convince other people to go along with our terribleness-decreasing government change proposals – it's important to keep on remembering the degree to which you can still pretty much do whatever you want.

In New York, Orthodox Jews with business disputes still bring them before a tribunal of rabbis, who judge them based on Jewish law. In Pennsylvania, the Amish live their own lives in their own way pretty much completely disconnected from US government decisions (although they needed a decent lobby group, the Amish Steering Committee, to work out a few special exemptions like from the draft). Socialists occasionally set up worker-owned companies run for the good of the proletariat, and they make products and earn money just like everyone else.

If you don't like the government, you're out of luck. But if you *and your whole community* don't like the government, you can organize your own internal relations however you want. You can't override existing laws – you'll still have to pay taxes, and you can't set up a bomb-making factory in your backyard. But you can add as many new laws as you want, enforced by threat of ostracism from your community, plus any other clever commitment mechanisms you can think of. There's nothing stopping communities – a broad term covering anything from villages to church congregations to cults to political organizations to online message boards – from creating internal welfare systems to help their poorer members, taking a say in when their members marry or divorce, making home schools that educated

their members' children, demanding their members in business treat their employees or business partners a certain way, et cetera.

Right now doctors' services are super-bloated and expensive because if a patient sues them they can be held liable for not filling out any of seven zillion forms or following any of twenty zillion best practices. But if the doctor only saw patients in their own community, and everyone in the community had mutual arbitration methods that worked better than the courts, maybe they could charge [a fraction](#) of the current price. This might not be illegal, as long as the community wasn't based on a protected group like race or religion. There just aren't many existing communities strong enough to make it work.

But some small seeds are starting to sprout. Social justice communities have sexual harassment policies much stronger than those of the country at large, and enforce them by ostracism and public shaming. Christians are trying to build the [Benedict Option](#), an embedded society that works on Christian norms and rules. And there's always the [seasteading movement](#), currently led by – oh, that's interesting – David Friedman's son. *Legal Systems Very Different From Ours* hints that we could build something like [Archipelago](#) gradually, without anybody noticing. The Jews and Gypsies did something like it. So did the Amish. Maybe all we have to do is start threatening to feud against each other's families, and utopia is right around the corner.

*
*
*

Legal Systems Very Different From Ours is available for free online at [this link](#).

Book Review: The Secret Of Our Success

Posted on June 4, 2019 by Scott Alexander



Previously in sequence: [Epistemic Learned Helplessness](#).

I

“Culture is the secret of humanity’s success” sounds like the most vapid possible thesis. [*The Secret Of Our Success*](#) by anthropologist Joseph Henrich manages to be an amazing book anyway.

Henrich wants to debunk (or at least clarify) a popular view where humans succeeded because of our raw intelligence. In this view, we are smart enough to invent neat tools that help us survive and adapt to unfamiliar environments.

Against such theories: we cannot actually do this. Henrich walks the reader through many stories about European explorers marooned in unfamiliar environments. These explorers usually starved to death. They starved to death in the middle of endless plenty. Some of them were in Arctic lands that the Inuit considered among their richest hunting grounds. Others were in jungles, surrounded by edible plants and animals. One particularly unfortunate group was in Alabama, and would have perished entirely if they hadn’t been captured and enslaved by local Indians first.

These explorers had many advantages over our hominid ancestors. For one thing, their exploration parties were made up entirely of strong young men in their prime, with no need to support women, children, or the elderly. They were often selected for their education and intelligence. Many of them were from Victorian Britain, one of the most successful civilizations in history, full of geniuses like Darwin and Galton. Most of them had some past experience with wilderness craft and survival. But despite their big brains, when faced with the task our big brains supposedly evolved for – figuring out how to do hunting and gathering in a wilderness environment – they failed pathetically.

Nor is it surprising that they failed. Hunting and gathering is actually really hard. Here’s Henrich’s description of how the Inuit hunt seals:

You first have to find their breathing holes in the ice. It’s important that the area around the hole be snow-covered—otherwise the seals will hear you and vanish. You then open the hole, smell it to verify it’s still in use (what do seals smell like?), and then assess the shape of the hole using a special curved piece of caribou antler. The hole is then covered with snow, save for a small gap at the top that is capped with a down indicator. If the seal enters the hole, the indicator moves, and you must blindly plunge your harpoon into the hole using all your weight. Your harpoon should be about 1.5 meters (5ft) long, with a detachable tip that is tethered with a heavy braid of sinew line. You can get the antler from the previously noted caribou, which you brought down with your driftwood bow.

The rear spike of the harpoon is made of extra-hard polar bear bone (yes, you also need to know how to kill polar bears; best to catch them napping in their dens). Once you’ve plunged your harpoon’s head into the seal, you’re then in a wrestling match as you reel him in, onto the ice, where you can finish him off with the aforementioned bear-bone spike.

Now you have a seal, but you have to cook it. However, there are no trees at this latitude for wood, and driftwood is too sparse and valuable to use routinely for fires. To have a reliable fire, you'll need to carve a lamp from soapstone (you know what soapstone looks like, right?), render some oil for the lamp from blubber, and make a wick out of a particular species of moss. You will also need water. The pack ice is frozen salt water, so using it for drinking will just make you dehydrate faster. However, old sea ice has lost most of its salt, so it can be melted to make potable water. Of course, you need to be able to locate and identify old sea ice by color and texture. To melt it, make sure you have enough oil for your soapstone lamp.

No surprise that stranded explorers couldn't figure all this out. It's more surprising that the Inuit *did*. And although the Arctic is an unusually hostile place for humans, Henrich makes it clear that hunting-gathering techniques of this level of complexity are standard everywhere. Here's how the Indians of Tierra del Fuego make arrows:

Among the Fuegians, making an arrow requires a 14-step procedure that involves using seven different tools to work six different materials. Here are some of the steps:

- The process begins by selecting the wood for the shaft, which preferably comes from chaura, a bushy, evergreen shrub. Though strong and light, this wood is a non-intuitive choice since the gnarled branches require extensive straightening (why not start with straighter branches?).
- The wood is heated, straightened with the craftsman's teeth, and eventually finished with a scraper. Then, using a pre-heated and grooved stone, the shaft is pressed into the grooves and rubbed back and forth, pressing it down with a piece of fox skin. The fox skin becomes impregnated with the dust, which prepares it for the polishing stage (Does it have to be fox skin?).
- Bits of pitch, gathered from the beach, are chewed and mixed with ash (What if you don't include the ash?).
- The mixture is then applied to both ends of a heated shaft, which must then be coated with white clay (what about red clay? Do you have to heat it?). This prepares the ends for the fletching and arrowhead.
- Two feathers are used for the fletching, preferably from upland geese (why not chicken feathers?).
- Right-handed bowman must use feathers from the left wing of the bird, and vice versa for lefties (Does this really matter?).
- The feathers are lashed to the shaft using sinews from the back of the guanaco, after they are smoothed and thinned with water and saliva (why not sinews from the fox that I had to kill for the aforementioned skin?).

Next is the arrowhead, which must be crafted and then attached to the shaft, and of course there is also the bow, quiver and archery skills. But, I'll leave it there, since I think you get the idea.

How do hunter-gatherers know how to do all this? We usually summarize it as "culture". How did it form? Not through some smart Inuit or Fuegian person reasoning it out; if that had been it, smart European explorers should have been able to reason it out too.

The obvious answer is "[cultural evolution](#)", but Henrich isn't much better than anyone else at taking the mystery out of this phrase. Trial and error must have been involved, and less successful groups/people imitating the techniques of more successful ones. But is that really a satisfying explanation?

I found the chapter on language a helpful reminder that we already basically accept something like this is true. How did language get invented? I'm especially interested in this question because of my brief interactions with conlanging communities – people who try to construct their own languages as a hobby or as part of a fantasy universe, like Tolkien did with Elvish. Most people are *terrible* at this; their languages are either unusable, or exact clones of English. Only people who (like Tolkien) already have years of formal training in linguistics can do a remotely passable job. And you're telling me the original languages were invented by cavemen? Surely there was no committee of Proto-Indo-European nomads that voted on whether to have an inflecting or agglutinating tongue? Surely nobody ran out of their cave shouting "Eureka!" after having discovered the interjection? We just kind of accept that after cavemen working really hard to communicate with each other, eventually language – still one of the most complicated and impressive productions of the human race – just sort of happened.

(this is how I feel about biological evolution too – how do you evolve an eye by trial and error? I've read papers speculating on the exact process, and they make lots of good points, but I still don't feel *happy* about it, like "Oh, of course this would happen!" At some point you just have to accept evolution is smarter than you are and smarter than you would expect to be possible.)

Taking the generation of culture as secondary to this kind of mysterious process, Henrich turns to its transmission. If cultural generation happens at a certain rate, then the fidelity of transmission determines whether a given society advances, stagnates, or declines.

For Henrich, humans started becoming more than just another species of monkey when we started transmitting culture with high fidelity. Some anthropologists talk about the [Machiavellian Intelligence Hypothesis](#) – the theory that humans evolved big brains in order to succeed at social maneuvering and climbing dominance hierarchies. Henrich counters with his own Cultural Intelligence Hypothesis – humans evolved big brains in order to be able to maintain things like Inuit seal hunting techniques. Everything that separates us from the apes is part of an evolutionary package designed to help us maintain this kind of culture, exploit this kind of culture, or adjust to the new abilities that this kind of culture gave us.

II

Secret gives many examples of many culture-related adaptations, and not all are in the brain.

Our digestive tracts evolved alongside our cultures. Specifically, they evolved to be unusually puny:

Our mouths are the size of the squirrel monkey's, a species that weighs less than three pounds. Chimpanzees can open their mouths twice as wide as we can and hold substantial amounts of food compressed between their lips and large teeth. We also have puny jaw muscles that reach up only to just below our ears. Other primates' jaw muscles stretch to the top of their heads, where they sometimes even latch onto a central bony ridge. Our stomachs are small, having only a third of the surface area that we'd expect for a primate of our size, and our colons are too short, being only 60% of their expected mass.

Compared to other animals, we have such atrophied digestive tracts that we shouldn't be able to live. What saves us? All of our food processing techniques, especially cooking, but also chopping, rinsing, boiling, and soaking. We've done much of the work of digestion before food even enters our mouths. Our culture teaches us how to do this, both in broad terms like "hold things over fire to cook them" and in specific terms like "this plant needs to be soaked in water for 24 hours to leach out the toxins". Each culture has its own cooking knowledge related to the local plants and animals; a frequent cause of death among European explorers was cooking things in ways that didn't unlock any of the nutrients, and so starving while apparently well-fed.

Fire is an especially important food processing innovation, and it is entirely culturally transmitted. Henrich is kind of cruel in his insistence on this. He recommends readers go outside and try to start a fire. He even gives some helpful hints – flint is involved, rubbing two sticks together works for some people, etc. He predicts – and stories I've heard from unfortunate campers confirm – that you will not be able to do this, despite an IQ far beyond that of most of our hominid ancestors. In fact, some groups (most notably the aboriginal Tasmanians) seem to have lost the ability to make fire, and never rediscovered it. Fire-making was discovered a small number of times, maybe once, and has been culturally transmitted since then.

But it's not just about chopping things up or roasting them. Traditional food processing techniques can get arbitrarily complicated. Nixtamalization of corn, necessary to prevent vitamin deficiencies, involves soaking the corn in a solution containing ground-up burnt seashells. The ancient Mexicans discovered this and lived off corn just fine for millennia. When the conquistadors took over, they ignored it and ate corn straight. For four hundred years, Europeans and Americans ate unnxxtamalized corn. By official statistics, three million Americans came down with corn-related vitamin deficiencies during this time, and up to a hundred thousand died. It wasn't until 1937 that Western scientists discovered which vitamins were involved and developed an industrial version of nixtamalization that made corn safe. Early 1900s Americans were very smart and had lots of advantages over ancient Mexicans. But the ancient Mexicans' culture got this one right in a way it took Westerners centuries to match.

Our hands and limbs also evolved alongside our cultures. We improved dramatically in some areas: after eons of tool use, our hands outclass those of any other ape in terms of finesse. In other cases, we devolved systems that were no longer necessary; we are much weaker than any other ape. Henrich describes a circus act of the 1940s where the ringmaster would challenge strong men in the audience to wrestle a juvenile chimpanzee. The chimpanzee was tied up, dressed in a mask that prevented it from biting, and wearing soft gloves that prevented it from scratching. No human ever lasted more than five seconds. Our common ancestor with other apes grew weaker and weaker as we became more and more reliant on artificial weapons to give us an advantage.

Even our sweat glands evolved alongside culture. Humans are persistence hunters: they cannot run as fast as gazelles, but they can keep running for longer than gazelles (or almost anything else). Why did we evolve into that niche? The secret is our ability to carry water. Every hunter-gatherer culture has invented its own water-carrying techniques, usually some kind of waterskin. This allowed humans to switch to perspiration-based cooling systems, which allowed them to run as long as they want.

III

But most of our differences from other apes are indeed in the brain. They're just not where you'd expect.

Tomasello et al tested human toddlers vs. apes on a series of traditional IQ type questions. The match-up was surprisingly fair; in areas like memory, logic, and spatial reasoning, the three species did about the same. But in ability to learn from another person, humans wiped the floor with the other two ape species:

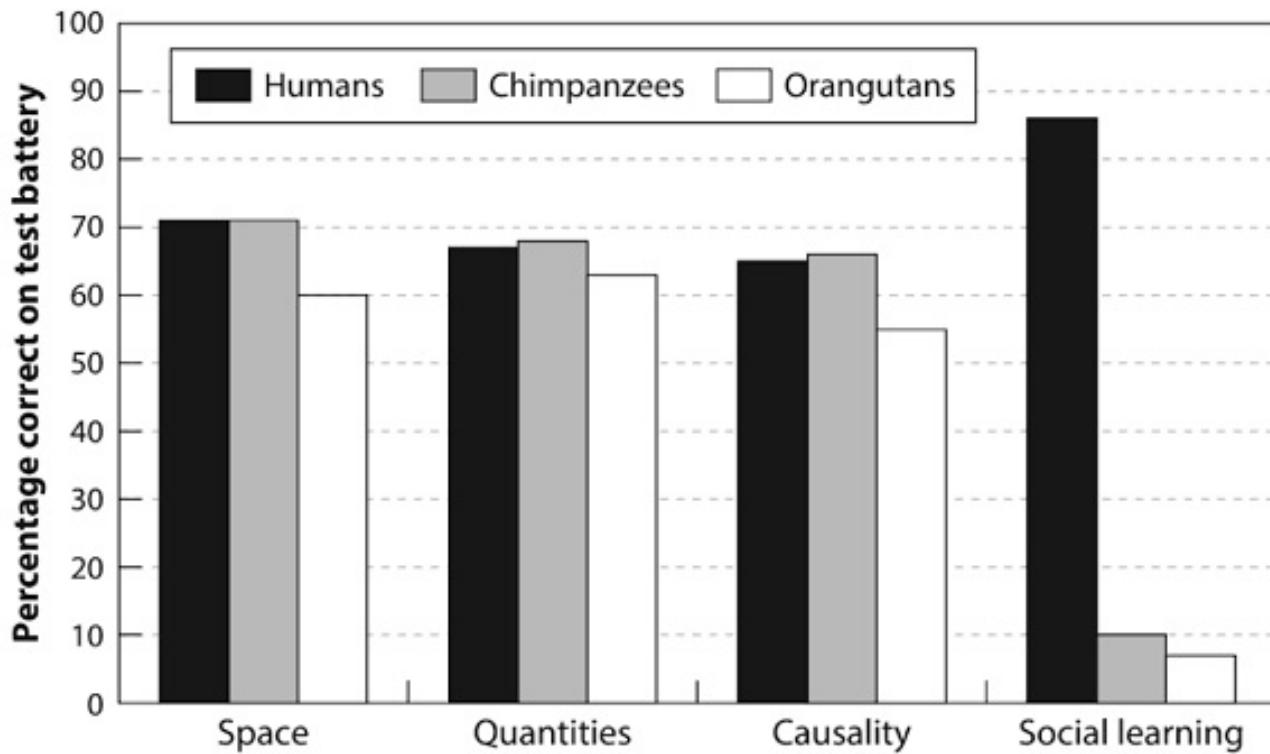


Figure 2.2. Average performance on four sets of cognitive tests with chimpanzees, orangutans, and toddlers.

Remember, Henrich thinks culture accumulates through random mutation. Humans don't have control over how culture gets generated. They have more control over how much of it gets transmitted to the next generation. If 100% gets transmitted, then as more and more mutations accumulate, the culture becomes better and better. If less than 100% gets transmitted, then at some point new culture gained and old culture lost fall into equilibrium, and your society stabilizes at some higher or lower technological level. This means that transmitting culture to the next generation is maybe the core human skill. The human brain is optimized to make this work as well as possible.

Human children are obsessed with learning things. And they don't learn things randomly. There seem to be "biases in cultural learning", ie slots in an infant's mind that they know need to be filled with knowledge, and which they preferentially seek out the knowledge necessary to fill.

One slot is for language. Human children naturally listen to speech (as early as in the womb). They naturally prune the phonemes they are able to produce and distinguish to the ones in the local language. And they naturally figure out how to speak and understand what people are saying, even though learning a language is hard even for smart adults.

Another slot is for animals. In a world where megafauna has been relegated to zoos, we *still* teach children their ABCs with "L is for lion" and "B is for bear", and children *still* read picture books about Mr. Frog and Mrs. Snake holding tea parties. Henrich suggests that just as the young brain is hard-coded to want to learn language, so it is hard-coded to want to learn the local animal life (maybe little boys' vehicle obsession is an outgrowth of this – buses and trains are the closest thing to local megafauna that most of them will encounter!)

Another slot is for plants:

To see this system in operation, let's consider how infants respond to unfamiliar plants. Plants are loaded

with prickly thorns, noxious oils, stinging nettles and dangerous toxins, all genetically evolved to prevent animals like us from messing with them. Given our species wide geographic range and diverse use of plants as foods, medicines and construction materials, we ought to be primed to both learn about plants and avoid their dangers. To explore this idea in the lab, the psychologists Annie Wertz and Karen Wynn first gave infants, who ranged in age from eight to eighteen months, an opportunity to touch novel plants (basil and parsley) and artifacts, including both novel objects and common ones, like wooden spoons and small lamps.

The results were striking. Regardless of age, many infants flatly refused to touch the plants at all. When they did touch them, they waited substantially longer than they did with the artifacts. By contrast, even with the novel objects, infants showed none of this reluctance. This suggests that well before one year of age infants can readily distinguish plants from other things, and are primed for caution with plants. But, how do they get past this conservative predisposition?

The answer is that infants keenly watch what other people do with plants, and are only inclined to touch or eat the plants that other people have touched or eaten. In fact, once they get the 'go ahead' via cultural learning, they are suddenly interested in eating plants. To explore this, Annie and Karen exposed infants to models who both picked fruit from plants and also picked fruit-like things from an artifact of similar size and shape to the plant. The models put both the fruit and the fruit-like things in their mouths. Next, the infants were given a choice to go for the fruit (picked from the plant) or the fruit-like things picked from the object. Over 75% of the time the infants went for the fruit, not the fruit-like things, since they'd gotten the 'go ahead' via cultural learning.

As a check, the infants were also exposed to models putting the fruit or fruit-like things behind their ears(not in their mouths). In this case, the infants went for the fruit or fruit-like things in equal measure. It seems that plants are most interesting if you can eat them, but only if you have some cultural learning cues that they aren't toxic.

After Annie first told me about her work while I was visiting Yale in 2013, I went home to test it on my 6-month-old son, Josh. Josh seemed very likely to overturn Annie's hard empirical work, since he immediately grasped anything you gave him and put it rapidly in his mouth. Comfortable in his mom's arms, I first offered Josh a novel plastic cube. He delighted in grabbing it and shoving it directly into his mouth, without any hesitation. Then, I offered him a sprig of arugula. He quickly grabbed it, but then paused, looked with curious uncertainty at it, and then slowly let it fall from his hand while turning to hug his mom.

It's worth pointing out how rich the psychology is here. Not only do infants have to recognize that plants are different from objects of similar size, shape and color, but they need to create categories for types of plants, like basil and parsley, and distinguish 'eating' from just 'touching'. It does them little good to code their observation of someone eating basil as 'plants are good to eat' since that might cause them to eat poisonous plants as well as basil. But, it also does them little good to narrowly code the observation as 'that particular sprig of basil is good to eat' since that particular sprig has just been eaten by the person they are watching. This another content bias in cultural learning.

This ties into the more general phenomenon of figuring out what's edible. Most Westerners learn insects aren't edible; some Asians learn that they are. This feels deeper than just someone telling you insects aren't edible and you believing them. When I was in Thailand, my guide offered me a giant cricket, telling me it was delicious. I believed him when he said it was safe to eat, I even believed him when he said it tasted good to him, but my conditioning won out – I didn't eat the cricket. There seems to be some process where a child's brain learns what is and isn't locally edible, then hard-codes it against future change.

(Or so they say; I've never been able to eat shrimp either.)

Another slot is for gender roles. By now we've all heard the stories of progressives who try to raise their children without any exposure to gender. Their failure has sometimes been taken as evidence that gender is hard-coded. But it can't be quite that simple: some modern gender roles, like girls = pink, are far from obvious or universal. Instead, it looks like children have a hard-coded slot that gender roles go into, work hard to figure out what the local gender roles are (even if their parents are trying to confuse them), then latch onto them and don't let go.

In the Cultural Intelligence Hypothesis, humans live in obligate symbiosis with a culture. A brain without an associated culture is incomplete and not very useful. So the infant brain is adapted to seek out the important aspects of its local culture almost from birth and fill them into the appropriate slots in order to become whole.

IV

The next part of the book discusses post-childhood learning. This plays an important role in hunter-gatherer tribes:

While hunters reach their peak strength and speed in their twenties, individual hunting success does not peak until around age 30, because success depends more on know-how and refined skills than on physical prowess.

This part of the book made most sense in the context of examples like the Inuit seal-hunting strategy which drove home just how complicated and difficult hunting-gathering was. Think less "Boy Scouts" and more "PhD"; a primitive tribesperson's life requires mastery of various complicated technologies and skills. And the difference between "mediocre hunter" and "great hunter" can be the difference between high status (and good mating opportunities) and low status, or even between life and death. Hunter-gatherers really want to learn the essentials of their hunter-gatherer lifestyle, and learning it is really hard. Their heuristics are:

Learn from people who are good at things and/or widely-respected. If you haven't already read about the difference between dominance and prestige hierarchies, check out [Kevin Simler's blog post](#) on the topic. People will fear and obey authority figures like kings and chieftains, but they give a different kind of respect ("prestige") to people who seem good at things. And since it's hard to figure out who's good at things (can a non-musician who wants to start learning music tell the difference between a merely good performer and one of the world's best?) most people use the heuristic of respecting the people who other people respect. Once you identify someone as respect-worthy, you strongly consider copying them in, well, everything:

To understand prestige as a social phenomenon, it's crucial to realize that it's often difficult to figure out what precisely makes someone successful. In modern societies, the success of a star NBA basketball player might arise from his:

1. intensive practice in the offseason
2. sneaker preference
3. sleep schedule
4. pre-game prayer
5. special vitamins
6. taste for carrots

Any or all of these might increase his success. A naïve learner can't tell all the causal links between an individual's practices and his success. As a consequence, learners often copy their chosen models broadly across many domains. Of course, learners may place more weight on domains that for one reason or other seem more causally relevant to the model's success. This copying often includes the model's personal habits or styles as well as their goals and motivations, since these may be linked to their success. This "if in doubt, copy it" heuristic is one of the reasons why success in one domain converts to influence across a broad range of domains.

The immense range of celebrity endorsements in modern societies shows the power of prestige. For example, NBA star Lebron James, who went directly from High School to the pros, gets paid millions to endorse State Farm Insurance. Though a stunning basketball talent, it's unclear why Mr. James is qualified to recommend insurance companies. Similarly, Michael Jordan famously wore Hanes underwear and apparently Tiger Woods drove Buicks. Beyonce' drinks Pepsi (at least in commercials). What's the connection between musical talent and sugary cola beverages?

Finally, while new medical findings and public educational campaigns only gradually influence women's approach to preventive medicine, Angelina Jolie's single OP-ED in the New York Times, describing her decision to get a preventive double mastectomy after learning she had the 'faulty' BRCA1 gene, flooded clinics from the U.K. to New Zealand with women seeking genetic screenings for breast cancer. Thus, an unwanted evolutionary side effect, prestige turns out to be worth millions, and represents a powerful and underutilized public health tool.

Of course, this creates the risk of prestige cascades, where some irrelevant factor (Henrich mentions being a reality show star) catapults someone to fame, everyone talks about them, and you end up with Muggeridge's definition of a celebrity: someone famous for being famous.

Some of this makes more sense if you go back to the evolutionary roots, and imagine watching the best hunter in your tribe to see what his secret is, or being nice to him in the hopes that he'll take you under his wing and teach you stuff.

(but if all this is true, shouldn't public awareness campaigns that hire celebrity spokespeople be wild successes? Don't they just as often fail, regardless of how famous a basketball player they can convince to lecture schoolchildren about how Winners Don't Do Drugs?)

Learn from people who are like you. If you are a man, it is probably a bad idea to learn fashion by observing women. If you are a servant, it is probably a bad idea to learn the rules of etiquette by observing how the king behaves. People are naturally inclined to learn from people more similar to themselves.

Henrich ties this in to various studies showing that black students learn best from a black teacher, female students from a female teacher, et cetera.

Learn from old people. Humans are almost unique in having menopause; most animals keep reproducing until they die in late middle-age. Why does evolution want humans to stick around without reproducing?

Because old people have already learned the local culture and can teach it to others. Henrich asks us to throw out any personal experience we have of elders; we live in a rapidly-changing world where an old person is probably "behind the times". But for most of history, change happened glacially slowly, and old people would have spent their entire lives accumulating relevant knowledge. Imagine a Silicon Valley programmer stumped by a particularly tough bug in his code calling up his grandfather, who has seventy years' experience in the relevant programming language.

Sometimes important events only happen once in a generation. Henrich tells the story of an Australian aboriginal tribe

facing a massive drought. Nobody knew what to do except Paralji, the tribe's oldest man, who had lived through the last massive drought and remembered where his own elders had told him to find the last-resort waterholes.

This same dynamic seems to play out even in other species:

In 1993, a severe drought hit Tanzania, resulting in the death of 20% of the African elephant calves in a population of about 200. This population contained 21 different families, each of which was led by a single matriarch. The 21 elephant families were divided into 3 clans, and each clan shared the same territory during the wet season (so, they knew each other). Researchers studying these elephants have analyzed the survival of the calves and found that families led by older matriarchs suffered fewer deaths of their calves during this drought.

Moreover, two of the three elephant clans unexpectedly left the park during the drought, presumably in search of water, and both had much higher survival rates than the one clan that stayed behind. It happens that these severe droughts only hit about once every four to five decades, and the last one hit about 1960. After that, sadly, elephant poaching in the 1970's killed off many of the elephants who would have been old enough in 1993 to recall the 1960 drought. However, it turns out that exactly one member of each of the two clans who left the park, and survived more effectively, were old enough to recall life in 1960. This suggests, that like Paralji in the Australian desert, they may have remembered what to do during a severe drought, and led their groups to the last water refuges. In the clan who stayed behind, the oldest member was born in 1960, and so was too young to have recalled the last major drought.

More generally, aging elephant matriarchs have a big impact on their families, as those led by older matriarchs do better at identifying and avoiding predators (lions and humans), avoiding internal conflicts and identifying the calls of their fellow elephants. For example, in one set of field experiments, researchers played lion roars from both male and female lions, and from either a single lion or a trio of lions. For elephants, male lions are much more dangerous than females, and of course, three lions are always worse than only one lion. All the elephants generally responded with more defensive preparations when they heard three lions vs. one. However, only the older matriarchs keenly recognized the increased dangers of male lions over female lions, and responded to the increased threat with elephant defensive maneuvers.

V

I was inspired to read *Secret* by [Scholar's Stage's review](#). I hate to be unoriginal, but after reading the whole book, I agree that the three sections Tanner cites – on divination, on manioc, and on shark taboos – are by far the best and most fascinating.

On divination:

When hunting caribou, Naskapi foragers in Labrador, Canada, had to decide where to go. Common sense might lead one to go where one had success before or to where friends or neighbors recently spotted caribou.

However, this situation is like [the [Matching Pennies](#) game]. The caribou are mismatchers and the hunters are matchers. That is, hunters want to match the locations of caribou while caribou want to mismatch the hunters, to avoid being shot and eaten. If a hunter shows any bias to return to previous spots, where he or others have seen caribou, then the caribou can benefit (survive better) by avoiding those locations (where they have previously seen humans). Thus, the best hunting strategy requires randomizing.

Can cultural evolution compensate for our cognitive inadequacies? Traditionally, Naskapi hunters decided where to go to hunt using divination and believed that the shoulder bones of caribou could point the way to success. To start the ritual, the shoulder blade was heated over hot coals in a way that caused patterns of cracks and burnt spots to form. This patterning was then read as a kind of map, which was held in a pre-specified orientation. The cracking patterns were (probably) essentially random from the point of view of hunting locations, since the outcomes depended on myriad details about the bone, fire, ambient temperature, and heating process. Thus, these divination rituals may have provided a crude randomizing device that helped hunters avoid their own decision-making biases.

This is not some obscure, isolated practice, and other cases of divination provide more evidence. In Indonesia, the Kantus of Kalimantan use bird augury to select locations for their agricultural plots. Geographer Michael Dove argues that two factors will cause farmers to make plot placements that are too risky. First, Kantu ecological models contain the Gambler's Fallacy, and lead them to expect floods to be less likely to occur in a specific location after a big flood in that location (which is not true). Second... Kantus pay attention to others' success and copy the choices of successful households, meaning that if one of their neighbors has a good yield in an area one year, many other people will want to plant there in the next year. To reduce the risks posed by these cognitive and decision-making biases, Kantu rely on a system of bird augury that effectively randomizes their choices for locating garden plots, which helps them avoid catastrophic crop failures. Divination results depend not only on seeing a particular bird species in a particular location, but also on what type of call the bird makes (one type of call may be favorable, and another unfavorable).

The patterning of bird augury supports the view that this is a cultural adaptation. The system seems to have evolved and spread throughout this region since the 17th century when rice cultivation was introduced. This makes sense, since it is rice cultivation that is most positively influenced by randomizing garden locations. It's possible that, with the introduction of rice, a few farmers began to use bird sightings as an indication of favorable garden sites. On-average, over a lifetime, these farmers would do better – be more successful – than farmers who relied on the Gambler's Fallacy or on copying others' immediate behavior. Whatever the process, within 400 years, the bird augury system spread throughout the agricultural populations of this Borneo region. Yet, it remains conspicuously missing or underdeveloped among local foraging groups and recent adopters of rice agriculture, as well as among populations in northern Borneo who rely on irrigation. So, bird augury has been systematically spreading in those regions where it's most adaptive.

Scott Aaronson has written about how easy it is to predict people trying to "be random":

In a class I taught at Berkeley, I did an experiment where I wrote a simple little program that would let people type either "f" or "d" and would predict which key they were going to push next. It's actually very easy to write a program that will make the right prediction about 70% of the time. Most people don't really know how to type randomly. They'll have too many alternations and so on. There will be all sorts of patterns, so you just have to build some sort of probabilistic model. Even a very crude one will do well. I couldn't even beat my own program, knowing exactly how it worked. I challenged people to try this and the program was getting between 70% and 80% prediction rates. Then, we found one student that the program predicted exactly 50% of the time. We asked him what his secret was and he responded that he "just used his free will."

But being genuinely random is important in pursuing mixed game theoretic strategies. Henrich's view is that divination solved this problem effectively.

I'm reminded of the Romans using augury to decide when and where to attack. This always struck me as crazy; generals are going to risk the lives of thousands of soldiers because they saw a weird bird earlier that morning? But war is a classic example of when a random strategy can be useful. If you're deciding whether to attack the enemy's right vs. left flank, it's important that the enemy can't predict your decision and send his best defenders there. If you're generally predictable – and Scott Aaronson says you are – then outsourcing your decision to weird birds might be the best way to go.

And then there's manioc. This is a tuber native to the Americas. It contains cyanide, and if you eat too much of it, you get cyanide poisoning. From Henrich:

In the Americas, where manioc was first domesticated, societies who have relied on bitter varieties for thousands of years show no evidence of chronic cyanide poisoning. In the Colombian Amazon, for example, indigenous Tukanoans use a multistep, multiday processing technique that involves scraping, grating, and finally washing the roots in order to separate the fiber, starch, and liquid. Once separated, the liquid is boiled into a beverage, but the fiber and starch must then sit for two more days, when they can then be baked and eaten. Figure 7.1 shows the percentage of cyanogenic content in the liquid, fiber, and starch remaining through each major step in this processing.

Such processing techniques are crucial for living in many parts of Amazonia, where other crops are difficult to cultivate and often unproductive. However, despite their utility, one person would have a difficult time figuring out the detoxification technique. Consider the situation from the point of view of the children and adolescents who are learning the techniques. They would have rarely, if ever, seen anyone get cyanide poisoning, because the techniques work. And even if the processing was ineffective, such that cases of goiter (swollen necks) or neurological problems were common, it would still be hard to recognize the link between these chronic health issues and eating manioc. Most people would have eaten manioc for years with no apparent effects. Low cyanogenic varieties are typically boiled, but boiling alone is insufficient to prevent the chronic conditions for bitter varieties. Boiling does, however, remove or reduce the bitter taste and prevent the acute symptoms (e.g., diarrhea, stomach troubles, and vomiting).

So, if one did the common-sense thing and just boiled the high-cyanogenic manioc, everything would seem fine. Since the multistep task of processing manioc is long, arduous, and boring, sticking with it is certainly non-intuitive. Tukanoan women spend about a quarter of their day detoxifying manioc, so this is a costly technique in the short term. Now consider what might result if a self-reliant Tukanoan mother decided to drop any seemingly unnecessary steps from the processing of her bitter manioc. She might critically examine the procedure handed down to her from earlier generations and conclude that the goal of the procedure is to remove the bitter taste. She might then experiment with alternative procedures by dropping some of the more labor-intensive or time-consuming steps. She'd find that with a shorter and much less labor-intensive process, she could remove the bitter taste. Adopting this easier protocol, she would have more time for other activities, like caring for her children. Of course, years or decades later her family would begin to develop the symptoms of chronic cyanide poisoning.

Thus, the unwillingness of this mother to take on faith the practices handed down to her from earlier generations would result in sickness and early death for members of her family. Individual learning does not pay here, and intuitions are misleading. The problem is that the steps in this procedure are causally opaque—an individual cannot readily infer their functions, interrelationships, or importance. The causal opacity of many cultural adaptations had a big impact on our psychology.

Wait. Maybe I'm wrong about manioc processing. Perhaps it's actually rather easy to individually figure out the detoxification steps for manioc? Fortunately, history has provided a test case. At the beginning of the seventeenth century, the Portuguese transported manioc from South America to West Africa for the first

time. They did not, however, transport the age-old indigenous processing protocols or the underlying commitment to using those techniques. Because it is easy to plant and provides high yields in infertile or drought-prone areas, manioc spread rapidly across Africa and became a staple food for many populations. The processing techniques, however, were not readily or consistently regenerated. Even after hundreds of years, chronic cyanide poisoning remains a serious health problem in Africa. Detailed studies of local preparation techniques show that high levels of cyanide often remain and that many individuals carry low levels of cyanide in their blood or urine, which haven't yet manifested in symptoms. In some places, there's no processing at all, or sometimes the processing actually increases the cyanogenic content. On the positive side, some African groups have in fact culturally evolved effective processing techniques, but these techniques are spreading only slowly.

Rationalists always wonder: how come people aren't more rational? How come you can prove a thousand times, using Facts and Logic, that something is stupid, and yet people will still keep doing it?

Henrich hints at an answer: for basically all of history, using reason would get you killed.

A reasonable person would have figured out there was no way for oracle-bones to accurately predict the future. They would have abandoned divination, failed at hunting, and maybe died of starvation.

A reasonable person would have asked why everyone was wasting so much time preparing manioc. When told "Because that's how we've always done it", they would have been unsatisfied with that answer. They would have done some experiments, and found that a simpler process of boiling it worked just as well. They would have saved lots of time, maybe converted all their friends to the new and easier method. Twenty years later, they would have gotten sick and died, in a way so causally distant from their decision to change manioc processing methods that nobody would ever have been able to link the two together.

Henrich discusses pregnancy taboos in Fiji; pregnant women are banned from eating sharks. Sure enough, these sharks contain chemicals that can cause birth defects. The women didn't really know why they weren't eating the sharks, but when anthropologists demanded a reason, they eventually decided it was because their babies would be born with shark skin rather than human skin. As explanations go, this leaves a lot to be desired. How come you can still eat other fish? Aren't you worried your kids will have scales? Doesn't the slightest familiarity with biology prove this mechanism is garbage? But if some smart independent-minded iconoclastic Fijian girl figured any of this out, she would break the taboo and her child would have birth defects.

In giving humans reason at all, evolution took a huge risk. Surely it must have wished there was some other way, some path that made us big-brained enough to understand tradition, but not big-brained enough to question it. Maybe it searched for a mind design like that and couldn't find one. So it was left with this ticking time-bomb, this ape that was constantly going to be able to convince itself of hare-brained and probably-fatal ideas.

Here, too, culture came to the rescue. One of the most important parts of any culture – more important than the techniques for hunting seals, more important than the techniques for processing tubers – is techniques for making sure nobody ever questions tradition. Like the belief that anyone who doesn't conform is probably a witch who should be cast out lest they bring destruction upon everybody. Or the belief in a God who has commanded certain specific weird dietary restrictions, and will torture you forever if you disagree. Or the fairy tales where the prince asks a wizard for help, and the wizard says "You may have everything you wish forever, but you must never nod your head at a badger", and then one day the prince nods his head at a badger, and his whole empire collapses into dust, and the moral of the story is that you should always obey weird advice you don't understand.

There's a monster at the end of this book. Humans evolved to transmit culture with high fidelity. And one of the biggest threats to transmitting culture with high fidelity was Reason. Our ancestors lived in Epistemic Hell, where they

had to constantly rely on causally opaque processes with justifications that couldn't possibly be true, and if they ever questioned them then they might die. Historically, Reason has been the villain of the human narrative, a corrosive force that tempts people away from adaptive behavior towards choices that "sounded good at the time".

Why are people so bad at reasoning? For the same reason they're so bad at letting poisonous spiders walk all over their face without freaking out. Both "skills" are really bad ideas, most of the people who tried them died in the process, so evolution removed those genes from the population, and successful cultures stigmatized them enough to give people an internalized fear of even trying.

VI

This book belongs alongside [*Seeing Like A State*](#) and the [works of G.K. Chesterton](#) as attempts to justify tradition, and to argue for organically-evolved institutions over top-down planning. What unique contribution does it make to this canon?

First, a lot more specifically anthropological / paleoanthropological rigor than the other two.

Second, a much crisper focus: Chesterton had only the fuzziest idea that he was writing about cultural evolution, and Scott was only a little clearer. I think Henrich is the only one of the three to use the term, and once you hear it, it's obviously the right framing.

Third, a sense of how traditions contain the meta-tradition of defending themselves against Reason, and a sense for why this is necessary.

And fourth, maybe we're not at the point where we really want unique contributions yet. Maybe we're still at the point where we have to have this hammered in by more and more examples. The temptation is always to say "Ah, yes, a few simple things like taboos against eating poisonous plants may be relics of cultural evolution, but obviously by now we're at the point where we know which traditions are important vs. random looniness, and we can rationally stick to the important ones while throwing out the garbage." And then somebody points out to you that *actually* divination using oracle bones was one of the important traditions, and if you thought you knew better than that and tried to throw it out, your civilization would falter.

Maybe we just need to keep reading more similarly-themed books until this point really sinks in, and we get properly worried.

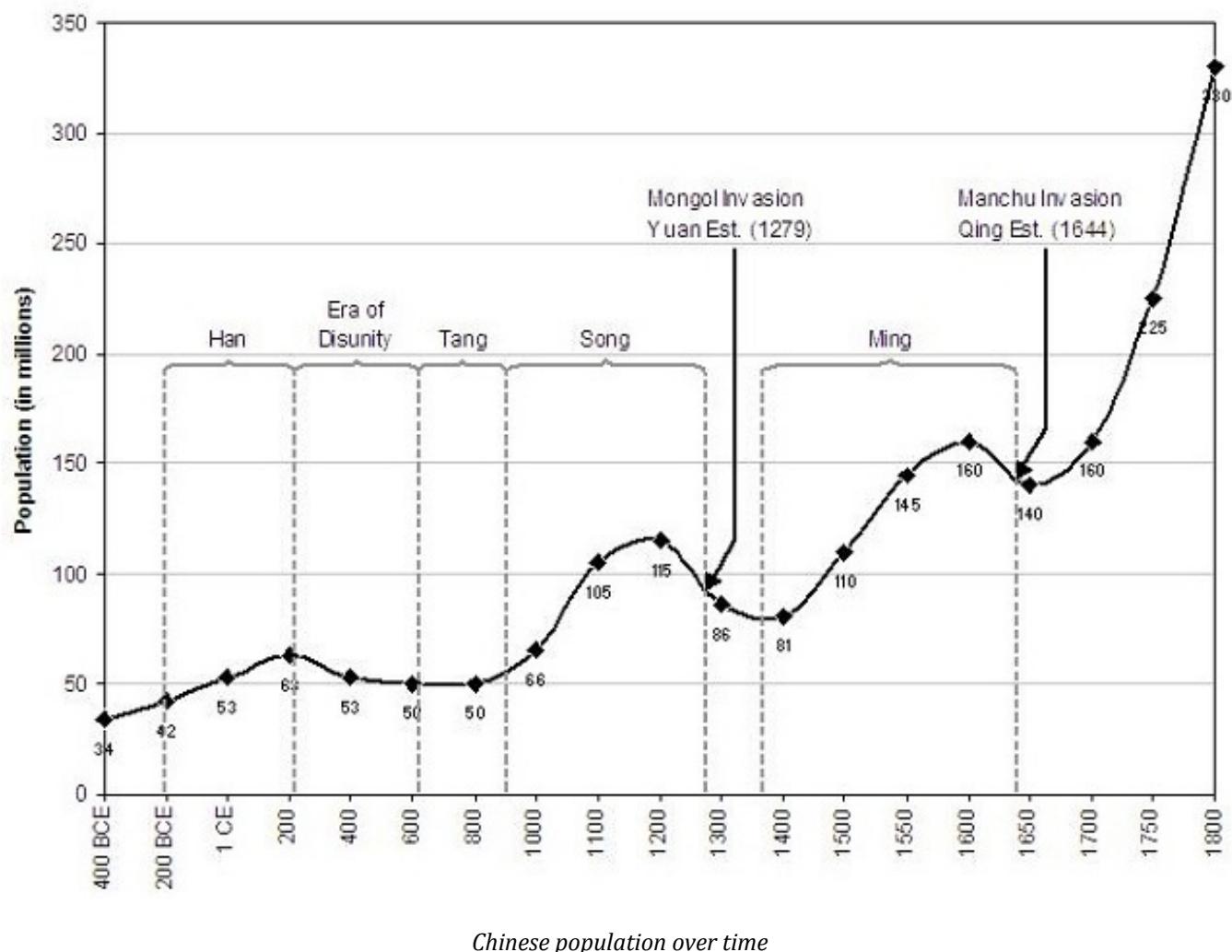
Book Review: Secular Cycles

Posted on August 12, 2019 by Scott Alexander



I

There is a tide in the affairs of men. It cycles with a period of about three hundred years. During its flood, farms and businesses prosper, and great empires enjoy golden ages. During its ebb, war and famine stalk the land, and states collapse into barbarism.



At least this is the thesis of Peter Turchin and Sergey Nefedov, authors of [Secular Cycles](#). They start off Malthusian: due to natural reproduction, population will keep increasing until it reaches the limits of what the land can support. At that point, everyone will be stuck at subsistence level. If any group ever enjoys a standard of living above subsistence level, they will keep reproducing until they are back down at subsistence.

Standard Malthusian theory evokes images of a population stable at subsistence level forever. But Turchin and Nefedov argues this isn't how it works. A population at subsistence will always be one meal away from starving. When a famine hits, many of them will starve. When a plague hits, they will already be too sickly to fight it off. When conflict arrives, they will be desperate enough to enlist in the armies of whichever warlord can offer them a warm meal.

These are not piecemeal events, picking off just enough of the population to bring it back to subsistence. They are great cataclysms. The Black Plague killed 30% – 60% of Europeans; the Antonine Plague of Rome was almost as deadly. The Thirty Years War killed 25% – 40% of Germans; the Time of Troubles may have killed 50% of medieval Russia.

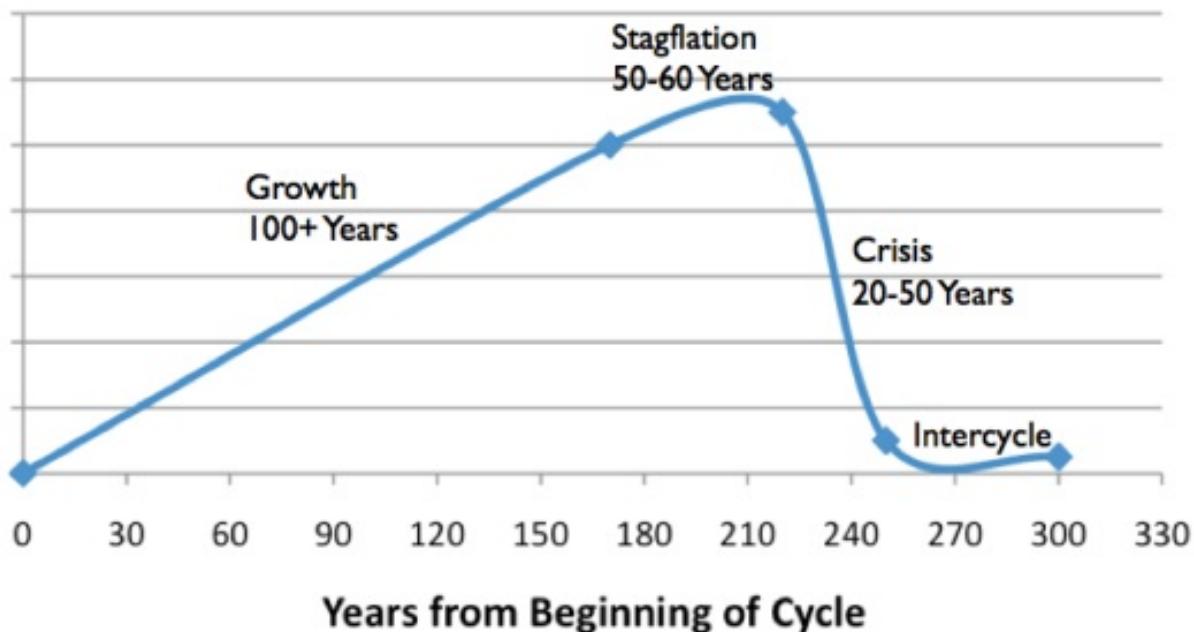
Thus the secular cycle. When population is low, everyone has more than enough land. People grow rich and reproduce. As time goes on, the same amount of farmland gets split among more and more people. Wages are driven down to subsistence. War, Famine, and Pestilence ravage the land, with Death not far behind. The killings continue until population is low again, at which point the cycle starts over.

This applies mostly to peasants, who are most at risk of starving. But nobles go through a related process. As a cycle begins, their numbers are low. As time goes on, their population expands, both through natural reproduction and through upward mobility. Eventually, there are more nobles than there are good positions...

(this part confused me a little. Shouldn't number of good positions scale with population? IE if one baron rules 1,000 peasants, the number of baronial positions should scale with the size of a society. I think T&N hint at a few answers. First, some positions are absolute rather than relative, eg "King" or "Minister of the Economy". Second, noble numbers may *sometimes* increase faster than peasant numbers, since nobles have more food and better chances to reproduce. Third, during boom times, the ranks of nobles are swelled through upward mobility. Fourth, conspicuous consumption is a ratchet effect: during boom times, the expectations of nobility should gradually rise. Fifth, sometimes the relevant denominator is not peasants but land: if a noble only has one acre of land, it doesn't matter how many peasants he controls. Sixth, nobles usually survive famines and plagues pretty well, so after those have done their work, there are far fewer peasants but basically the same number of nobles. All of these factors contribute to excess noble population – or as T&N call it, "elite overproduction")

...and the nobles form "rival patronage networks" to fight for the few remaining good spots. The state goes from united (or at least all nobles united against the peasants) to divided, with coalitions of nobles duking it out (no pun intended). This can lead either to successful peasant rebellion, as some nobles support the peasants as part of inter-noble power plays, or just to civil war. Although famine and plague barely affect nobles, war affects them disproportionately – both because they were often knights or other front-line soldiers, and because killing the other side's nobles was often a major strategic goal (think Game of Thrones). So a civil war usually further depletes the already-depleted peasant population, and finally depletes noble populations, leading to a general underpopulation and the beginning of the next cycle.

Shape of Typical "Secular Cycle"



Combine these two processes, and you get the basic structure of a secular cycle. There are about a hundred years of unalloyed growth, as peasant and noble populations rebound from the last disaster. During this period, the economy is strong, the people are optimistic and patriotic, and the state is strong and united.

After this come about fifty years of "stagflation". There is no more room for easy growth, but the system is able to absorb the surplus population without cracking. Peasants may not have enough land, but they go to the city in search of jobs. Nobles may not have enough of the positions they want, but they go to college in order to become bureaucrats, or join the retinues of stronger nobles. The price of labor reaches its lowest point, and the haves are able to exploit the desperation of the have-nots to reach the zenith of their power. From the outside, this period can look like a golden age: huge cities buzzing with people, universities crammed with students, ultra-rich nobles throwing money at the arts and sciences. From the inside, for most people it will look like a narrowing of opportunity and a hard-to-explain but growing sense that something is wrong.

After this comes a crisis. The mechanisms that have previously absorbed surplus population fail. Famine and disease ravage the peasantry. State finances fall apart. Social trust and patriotism disappear as it becomes increasingly obvious that it's every man for himself and that people with scruples will be defeated or exploited by people without.

After this comes the depression period (marked "intercycle" on the graph above, but I'm going to stick with the book's term). The graph makes it look puny, but it can last 100 to 150 years. During this period, the peasant population is low, but the noble population is still high. This is most likely an era of very weak or even absent state power, barbarian invasions, and civil war. The peasant population is in a good position to expand, but cannot do so because wars keep killing people off or forcing them into walled towns where they can't do any farming. Usually it takes a couple more wars and disasters before the noble population has decreased enough to reverse elite overproduction. At this point the remaining nobles look around, decide that there is more than enough for all of them, and feel incentivized to cooperate with the formation of a strong centralized state.

This cycle is interwoven with a second 40-60 year process that T&N call the "fathers-and-sons cycle" or

“bigenerational cycle”. The data tend to show waves of disorder about every 40-60 years. During the “integrative trend” (T&N’s term for the optimistic growth and stagflation phases), these can just be minor protests or a small rebellion that is easily crushed. During the “disintegrative trend” (crisis + depression), they usually represent individual outbreaks of civil war. For example, during the Roman Republic, the violence around the death of Tiberius Gracchus in 133 BC was relatively limited, because Rome had not yet entered its crisis phase. 40 years later, in the depths of the crisis phase, there was a second outbreak of violence (91 – 82 BC) including the Social War and Sulla’s wars, which escalated to full-scale (though limited) civil war. 40 years later there was a third outbreak (49 – 27 BC) including Caesar and Augustus’s very large civil wars. After that the new integrative trend started and further violence was suppressed.

In *Secular Cycles*, T&N mostly just identify this pattern from the data and don’t talk a lot about what causes it. But in some of Turchin’s other work, he applies some of the math used to model epidemics in public health. His model imagines three kinds of people: natives, radicals, and moderates. At the start of a cycle, most people are naive, with a few radicals. Radicals gradually spread radicalism, either by converting their friends or provoking their enemies (eg a terrorist attack by one side convinces previously disengaged people to join the other side). This spreads like any other epidemic. But as violence gets worse, some people convert to “moderates”, here meaning not “w wishy-washy people who don’t care” but something more like “people disenchanted with the cycle of violence, determined to get peace at any price”. Moderates suppress radicals, but as they die off most people are naive and the cycle begins again. Using various parameters for his model Turchin claims this predicts the forty-to-sixty year cycle of violence observed in the data.

So this is the basic thesis of *Secular Cycles*. Pre-industrial history operates on two cycles: first, a three-hundred year cycle of the rise-and-fall of civilizations. And second, a 40-60 year cycle of violent disorder that only becomes relevant during the lowest parts of the first cycle.

II

This is all in the first chapter of the book! The next eight chapters are case studies of eight different historical periods and how they followed the secular cycle model.

For example, Chapter 7 is on the Roman Empire. It starts with Augustus in 27 BC. The Roman Republic has just undergone a hundred years of civil war, from the Gracchi to Marius to Sulla to Pompey to Caesar to Antony. All of this decreased its population by 30% from its second-century peak. That means things are set to get a lot better very quickly.

The expansion phase of the Empire lasted from Augustus (27 BC) to Nerva (96 AD), followed by a stagflation phase from Nerva to Antonius Pius (165 AD). Throughout both phases, the population grew – from about 40 million in Augustus’ day to 65 million in Antonius’. Wheat prices stayed stable until Nerva, then doubled from the beginning of the second century to its end. Legionary pay followed the inverse pattern, staying stable until Nerva and then decreasing by a third before 200. The finances of the state were the same – pretty good until the late second century (despite occasional crazy people becoming Emperor and spending the entire treasury building statues of themselves), but cratering during the time of Marcus Aurelius and Commodus (who debased the denarius down to only 2 g silver).

Throughout expansion and stagflation, the Empire was relatively peaceful (the “Pax Romana”). Sure, occasionally a crazy person would become Emperor and they would have to kill him. There was even one small civil war which lasted all of a year (69 AD). But in general, these were isolated incidents.

Throughout the expansion phase, upward mobility was high and income inequality relatively low. T&N measure this as how many consuls (the highest position in the Roman governmental hierarchy) had fathers who were also consuls.

This decreased throughout the first century – from 46% to 18% – then started creeping back up during the stagflation phase to reach 32% at the end of the second century.

The crisis phase began in 165 AD at the peak of Rome's population and wealth. The Antonine Plague ravaged the Empire, killing 30% of the population. Fifteen years later, the century-long dominance of the Good Emperors ended, and Commodus took the throne. Then he was murdered and Pertinax took the throne. Then he was murdered and Didius Julianus took the throne. Then he was murdered and Septimius Severus took the throne.

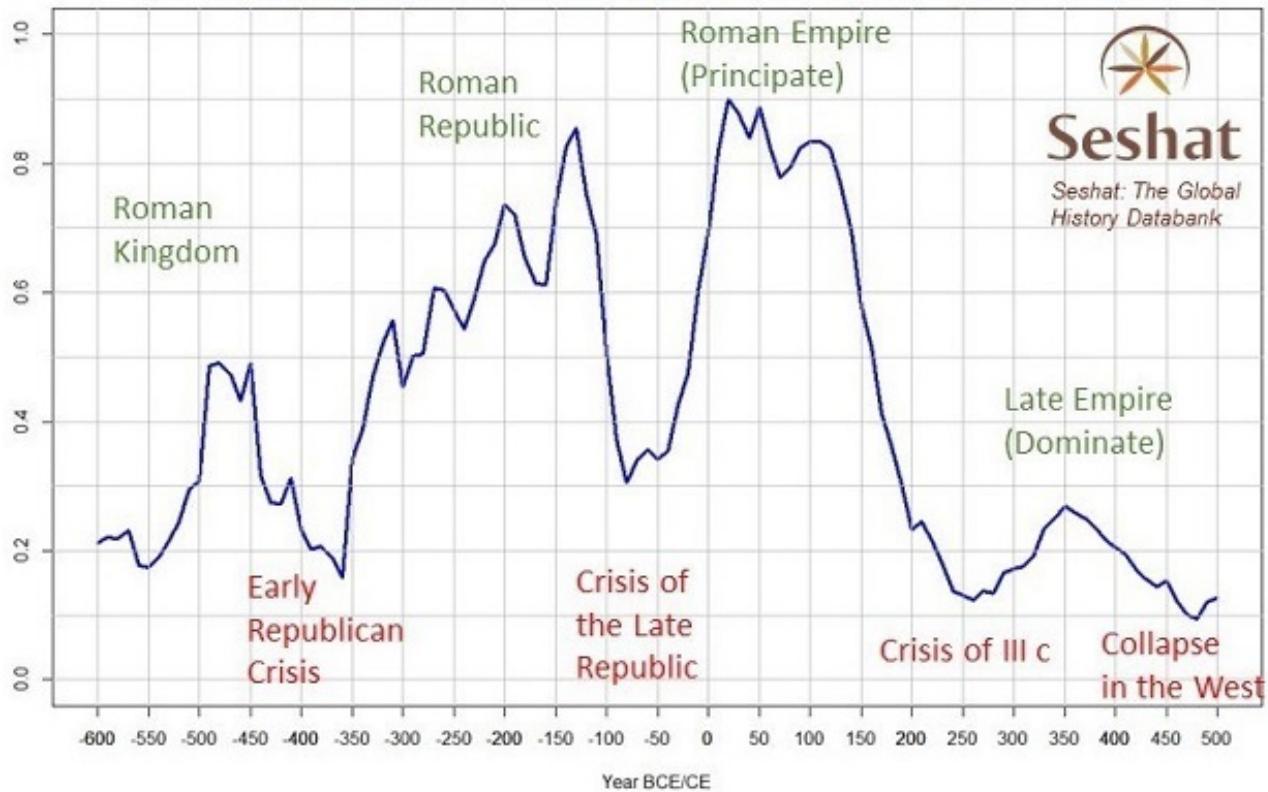
Now we are well into the disintegrative trend, and the shorter 40-60 year cycle comes into play. Septimius Severus founded a dynasty that lasts 41 years, until Septimius Alexander (the grandson of Septimius Severus' sister-in-law; it's complicated) was assassinated by his own soldiers in Germany. This begins the [Crisis Of The Third Century](#), a time of constant civil war, mass depopulation, and economic collapse. The Five Good Emperors of the second century ruled 84 years between them (average of 17 years per emperor). The fifty year Crisis included 27 emperors, for an average of less than 2 years per emperor.

Finally, in 284, Emperor Diocletian ended the civil wars, re-established centralized authority, and essentially refounded the Roman Empire – a nice round 310 years after Augustus did the same. T&N mark this as the end of a secular cycle and the beginning of a new integrative trend.

T&N are able to tell this story. But they don't just tell the story. They are able to cite various statistics to back themselves up. The Roman population statistics. The price of wheat and other foodstuffs. The average wages for laborers. They especially like coin hoards – the amount of buried treasure from a given period discovered by treasure-hunters – because they argue you only bury your money during times of instability, so this forms a semi-objective way of measuring how unstable things are.

They are at their best when presenting very broad summary statistics. For example, Roman industry produced vast amounts of lead, which entered the atmosphere and settled into the Greenland ice sheet. Here is Roman lead output per year as measured in ice cores:

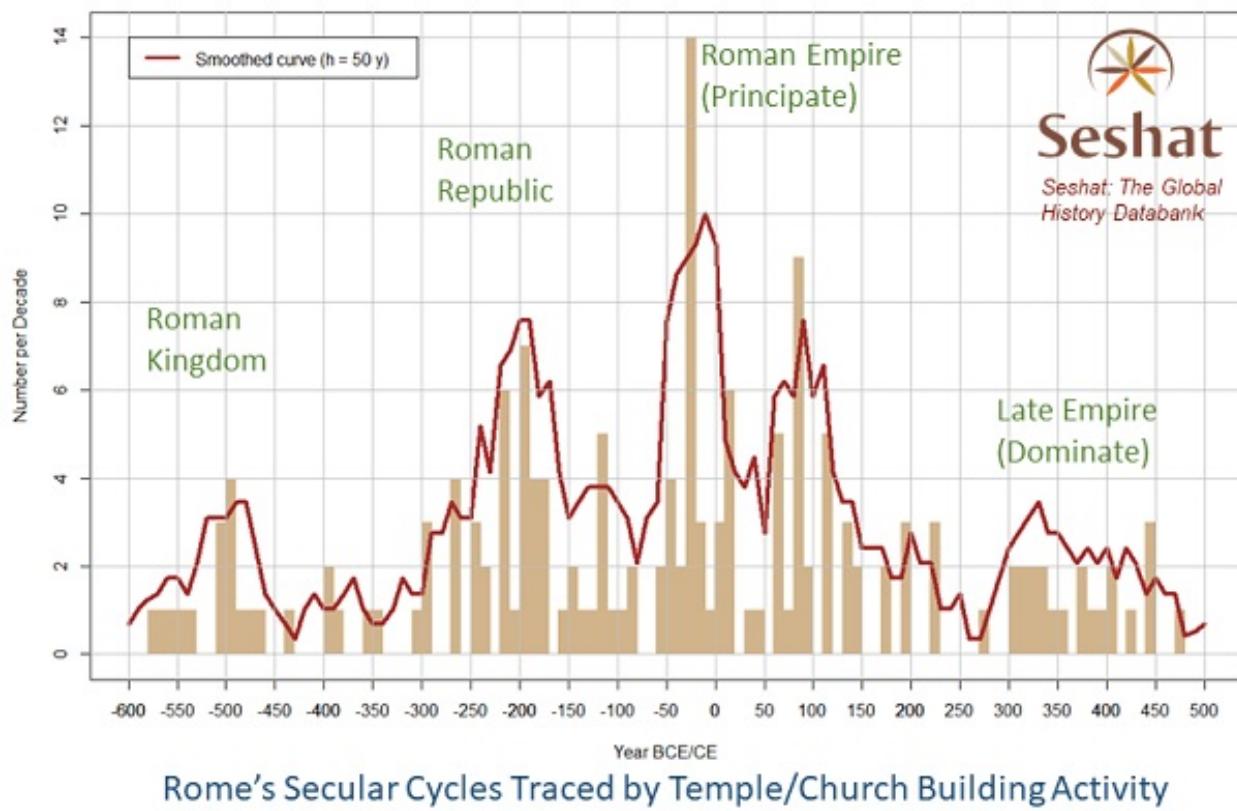
Rome's Secular Cycles Traced by Lead Pollution in Greenland Ice



This shows four peaks for the four cycles T&N identify in Rome: the Kingdom, the Republic, the Early Empire of Augustus (the one described above), and the Late Empire of Diocletian (Dominate). It even shows a sawtooth-y pattern corresponding to the shorter bigenerational cycles.

Or here is building activity in Rome, measured by how many buildings archaeologists have found from a given time:

Religious Buildings in Rome

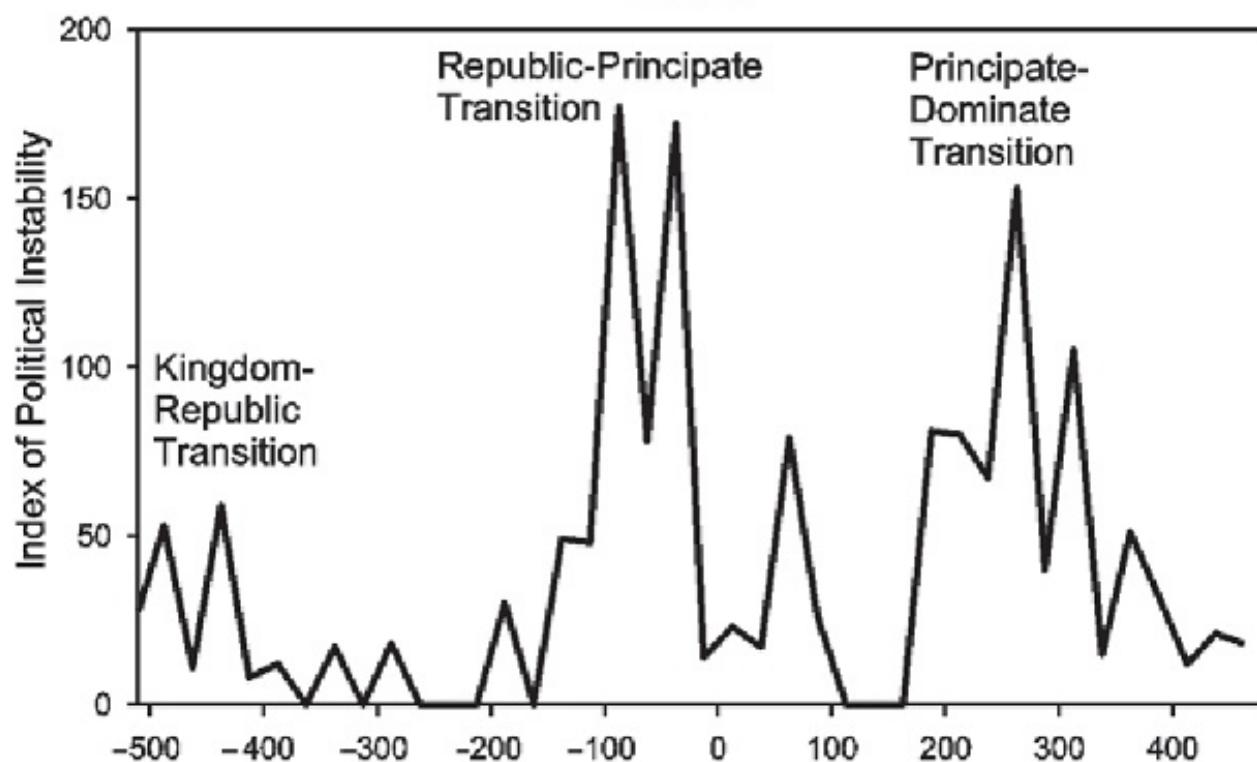


This is a little less perfect (why is there a big gap in the middle of the Principate? I guess Augustus is a hard act to follow, building-wise) but it still looks good for the cycle theory.

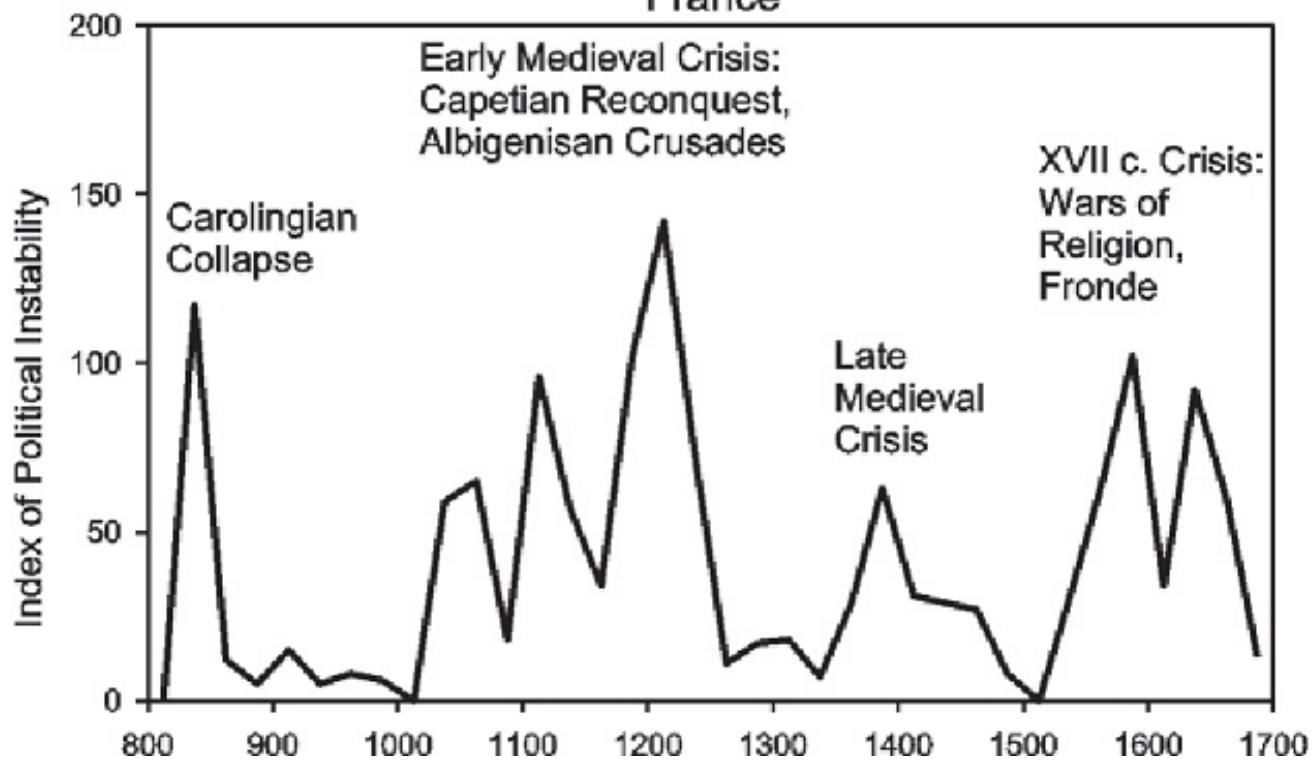
And here is an Index Of Political Instability, which “combines measures of duration, intensity, and scale of political instability events, coded by a team of professional historians”:

a

Rome

*b*

France



Rome is the one on top. Instability clearly peaks during the crisis-depression phases between T&N's secular cycles – again with a sawtooth pattern representing the bigenerational cycles.

III

Seeing [patterns in random noise](#) is one of the basic human failure modes. *Secular Cycles* is so *prima facie* crackpottish that it should require mountains of data before we even start wondering if it *might* be true. I want to make it clear that the book – plus Turchin individually in some of his other books and papers – provides these mountains. I can't show every single case study, graph, and table in this book review. But the chapter above on the Roman Principate included 25 named figures and graphs, plus countless more informal presentations of data series, from “real wages of agricultural laborers in Roman Egypt during the second century” to “mean annual real land rents for wheat fields in artabas per aroura, 27 BC to 268 CE” to “imperial handouts per reign-year” to “importation of African red slip ware into the Albegna Valley of Etruria, 100 – 600”. And this is just one chapter, randomly chosen. There are seven others just like this. This book understands the burden of proof it is under, and does everything it can to meet it.

Still, we should be skeptical. How many [degrees of freedom](#) do T&N have, and is it enough to undermine their case?

First, they get some freedom in the civilizations they use as case studies. They could have searched through every region and period and cherry-picked eight civilizations that rose and fell over a periods of three hundred years. Did they? I don't think so. The case studies are England, France, Rome, and Russia. These are some of the civilizations of greatest interest to the English-speaking world (except Russia, which makes sense in context because the authors are both Russian). They're also some of the civilizations best-studied by Anglophone historians and with the most data available (the authors' methodology requires having good time-series of populations, budgets, food production, etc).

Also, it's not too hard to look at the civilizations they didn't study and fill in the gaps. The book barely mentions China, but it seems to fit the model pretty well (“the empire united longs to divide; divided longs to unite”). In fact, taking the quotation completely seriously – the empire was first united during the Qin Dynasty starting in 221 BC, which lasted only 20 years before seguing into the Han Dynasty in 202 BC. The Han expanded and prospered for about a century, had another century of complicated intrigue and frequently revolt, and then ended in disaster in the first part of the first century, with a set of failed reforms, civil war, the sack of the capital, some more civil war, peasant revolt, and even more civil war. The separate period of the Eastern Han Dynasty began in 25 AD, about 240 years after the beginning of the Qin-Han cycle. The Eastern Han also grew and prospered for about a hundred years, then had another fifty years of simmering discontent, then fell apart in about 184 AD, with another series of civil wars, peasant rebellions, etc. This was the Three Kingdoms Period during which “the empire united longs to divide, divided longs to unite” was written to describe. It lasted another eighty years until 266 AD, after which the Jin Dynasty began. The Jin Dynasty was kind of crap, but it lasted another 180 years until 420, followed by 160 years of division, followed by the Sui and Tang dynasties, which were not crap. So I don't think it takes too much pattern-matching to identify a Western-Han-to-Eastern-Han Cycle of 240 years, followed by an Eastern-Han-to-Jin Cycle of 241 years, followed by a Jin-to-Sui/Tang-Cycle of 324 years.

One could make a more hostile analysis. Is it really fair to lump the Western Jin and Eastern Jin conveniently together, but separate the Western Han and Eastern Han conveniently apart? Is it really fair to call the crappy and revolt-prone Jin Dynasty an “integrative trend” rather than a disintegrative trend that lasted much longer than the theory should predict? Is it really fair to round off cycles of 240 and 320 years to “basically 300 years”?

I think the answer to all of these is “T&N aren't making predictions about the length of Chinese dynasties, they're making predictions about the nature of secular cycles, which are correlated with dynasties but not identical to them”. If I had the equivalent to lead core readings for China, or an “instability index”, or time series data for wages or health or pottery importation or so on, maybe it would be perfectly obvious that the Eastern and Western Han defined two different periods, but the Eastern and Western Jin were part of the same period – the same way one look at the lead core data for Rome shows that the Julio-Claudian dynasty vs. the Flavian Dynasty is not an interesting transition.

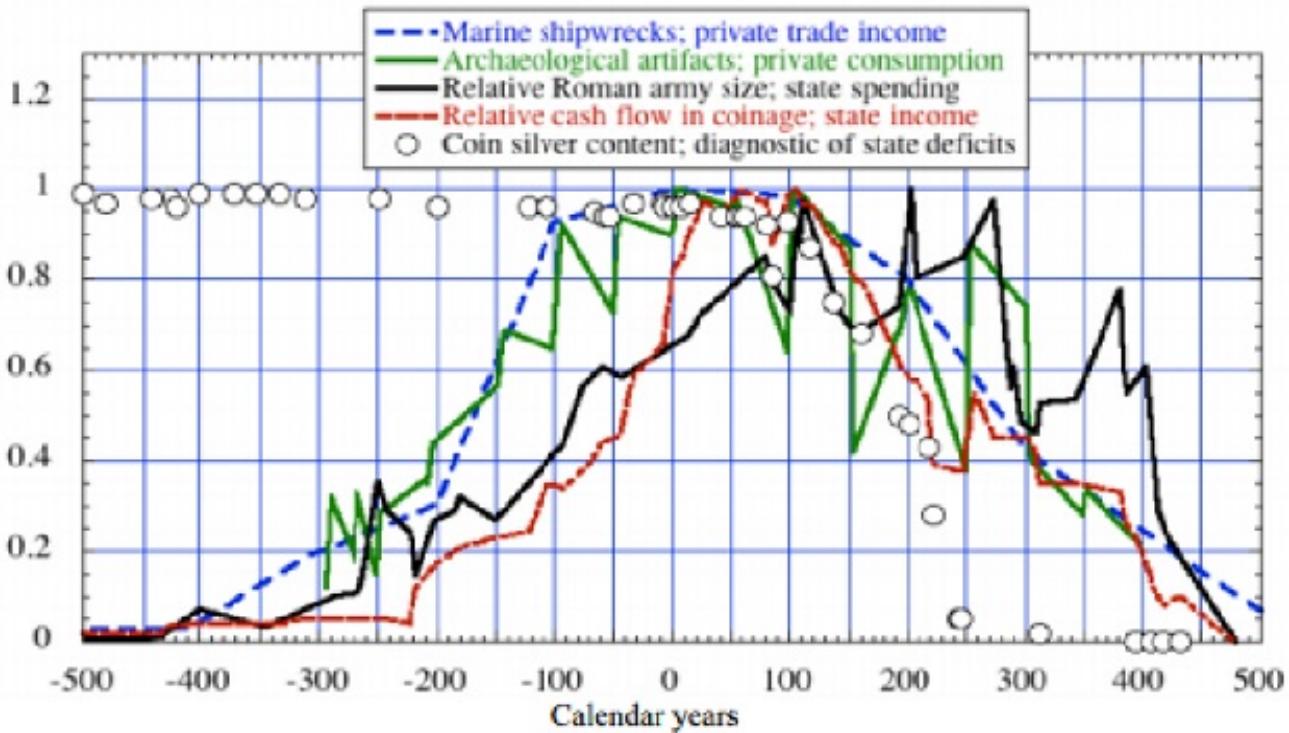
A secondary answer might be that T&N admit all sorts of things can alter the length of secular cycles. They tragically

devote only a few pages to “Ibn Khaldun cycles”, the theory of 14th century Arabic historian Ibn Khaldun that civilizations in the Maghreb rise and fall on a one hundred year period. But they discuss it just enough to say their data confirm Ibn Khaldun’s observations. The accelerated timescale (100 vs. 300 years) is because the Maghreb is massively polygynous, with successful leaders having harems of hundreds of concubines. This speeds up the elite overproduction process and makes everything happen in fast-forward. T&N also admit that their theory only describes civilizations insofar as they are self-contained. This approximately holds for hegemons like Rome at its height, but fails for eg Poland, whose history is going to be much more influenced by when Russia or Germany decides to invade than by the internal mechanisms of Polish society. Insofar as external shocks – whether climatic, foreign military, or whatever else – affect a civilization, secular cycles will be stretched out, compressed, or just totally absent.

This sort of thing must obviously be true, and it’s good T&N say it, but it’s also a free pass to add as many epicycles as you need to explain failure to match data. All I can say looking at China is that, if you give it some wiggle room, it seems to fit T&N’s theories okay. The same is true of a bunch of other civilizations I plugged in to see if they would work.

Second, T&N get some degrees of freedom based on what statistics they use. In every case, they present statistics that support the presence of secular cycles, but they’re not the same statistics in every case. On the one hand, this is unavoidable; we may not have good wage data for every civilization, and levels of pottery importation might be more relevant to ancient Rome than to 19th-century Russia. On the other hand, I’m not sure what prevents them from just never mentioning the Instability Index if the Instability Index doesn’t show what they want it to show.

Here are some random Rome-related indicators I found online:



None of them show the same four-peaked Kingdom-Republic-Principate-Dominate pattern as the ones *Secular Cycles* cites, or the ones Turchin has online.

Third, a lot of the statistics themselves have some degrees of freedom. A lot of them are things like “Instability Index” or “Index of Social Well-Being” or “General Badness Index”. These seem like the kind of scores you can fiddle with to get the results you want. Turchin claims he hasn’t fiddled with them – his instability index is taken from a 1937 paper I

haven't been able to find. But how many papers like that are there? Am I getting too conspiratorial now?

Likewise, we don't have direct access to the budget of the Roman Empire (or Plantagenet England, or...). Historians have tried to reconstruct it based on archaeology and the few records that have survived. T&N cite these people, and the people they cite are at the top of their fields and say what T&N say they say. But how much flexibility did they have in deciding which estimate of the Roman budget to cite? Is there enough disagreement that they could cite the high estimate for one period and the low estimate for another, then prove it had gone down? I don't know (though a few hours' work ought to be enough to establish this).

I wish I could find commentary by other academics and historians on *Secular Cycles*, or on Turchin's work more generally. I feel like somebody should either be angrily debunking this, or else throwing the authors a ticker-tape parade for having solved history. Neither is happening. The few comments I can find are mostly limited to navel gazing about whether history should be quantitative or qualitative. The few exceptions find are [blog posts by people I already know and respect urging me](#) to read Turchin five years ago, advice I am sorry for not taking. If you know of any good criticism, please tell me where to find it.

Until then, my very quick double-checking suggests T&N are pretty much on the level. But there could still be subtler forms of overfitting going on that I don't know enough about history to detect.

IV

If this is true, does it have any implications for people today?

First, a very weak implication: it makes history easier to learn. I was shocked how much more I remembered about the Plantagenets, Tudors, Capetians, etc after reading this book, compared to reading any normal history book about them. I think the secret ingredient is structure. If history is just "one damn thing after another", there's no framework for figuring out what matters, what's worth learning, what follows what else. The secular cycle idea creates a structure that everything fits into neatly. I know that the Plantagenet Dynasty lasted from 1154 – 1485, because it *had to*, because that's a 331 year secular cycle. I know that the important events to remember include the Anarchy of 1135 – 1153 and the War of the Roses from 1455 – 1487, because those are the two crisis-depression periods that frame the cycle. I know that after 1485 Henry Tudor took the throne and began a new age of English history, because that's the beginning of the integrative phase of the next cycle. All of this is a lot easier than trying to remember these names and dates absent any context. I would recommend this book for that reason alone.

Second, I think this might give new context to Piketty on inequality. T&N describe inequality as starting out very low during the growth phase of a secular cycle, rising to a peak during the stagflation phase, then dropping precipitously during the crisis. Piketty describes the same: inequality rising through the peaceful period of 1800 to 1900, dropping precipitously during the two World Wars, then gradually rising again since then. This doesn't make a huge amount of sense, since I'm not sure you can fit the post industrial world into secular cycles. But I notice Piketty seems to think of this as a once-off event – inequality has been rising forever, broken only by the freak crisis of the two World Wars – and it's interesting to read T&N talk about the exact same process recurring again and again throughout history.

Finally, and most important: is there any sense in which this is still going on?

The easiest answer would be no, there isn't. The secular cycles are based around Malthusian population growth, but we are now in a post-Malthusian regime where land is no longer the limiting resource. And the cycles seem to assume huge crises killing off 30% to 50% of the population, but those don't happen anymore in First World countries; the Civil War was the bloodiest period of US history, and even it only killed 2% of Americans. Even Germany only lost about 15% of its population in World Wars I + II.

But Turchin has another book, *Ages Of Discord*, arguing that they do. I have bought it and started it and will report back when I'm done.

Even without a framework, this is just interesting to think about. In popular understanding of American history, you can trace out optimistic and pessimistic periods. The national narrative seems to include a story of the 1950s as a golden age of optimism. Then everyone got angry and violent in the early 1970s (the Status 451 review of [Days Of Rage](#) is pretty great here, and reminds us that “people have completely forgotten that in 1972 we had over nineteen hundred domestic bombings in the United States”). Then everything suddenly got better once Reagan declared “morning in America” in the 1980s, with an era of optimism and good feelings lasting through the Clinton administration. Then things starting to turn bad sometime around Bush II. And now everybody hates each other, and fascists and antifa are fighting in the streets, and people are talking about how “civility” and “bipartisanship” are evil tools of oppression, and PredictIt says an avowed socialist has a 10% chance of becoming president of the US. To what extent is this narrative true? I don't know, but it's definitely the narrative.

One thing that strikes me about T&N's cycles is the ideological component. They describe how, during a growth phase, everyone is optimistic and patriotic, secure in the knowledge that there is enough for everybody. During the stagflation phase, inequality increases, but *concern about* inequality increases even more, zero-sum thinking predominates, and social trust craters (both because people are actually defecting, and because it's in lots of people's interest to play up the degree to which people are defecting). By the crisis phase, partisanship is much stronger than patriotism and radicals are talking openly about how violence is ethically obligatory.

And then, eventually, things get better. There is a new Augustan Age of virtue and the reestablishment of all good things. This is a really interesting claim. Western philosophy tends to think in terms of trends, not cycles. We see everything going on around us, and we think this is some endless trend towards more partisanship, more inequality, more hatred, and more state dysfunction. But *Secular Cycles* offers a narrative where endless trends can end, and things can get better after all.

Of course, it also offers a narrative where sometimes this process involves the death of 30% – 50% of the population. Maybe I should read Turchin's other books before speculating any further.

Book Review: Ages Of Discord

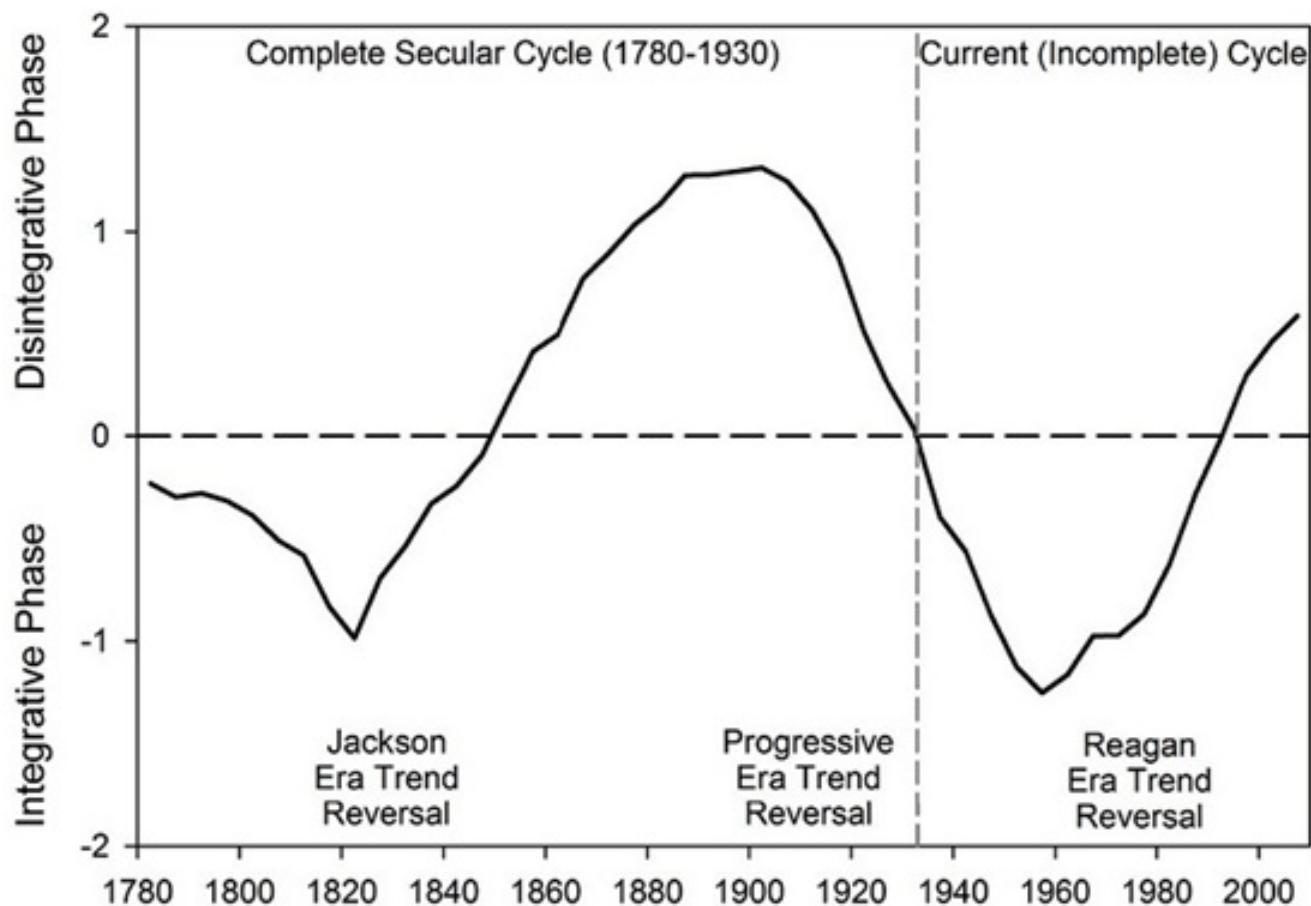
Posted on September 2, 2019 by Scott Alexander



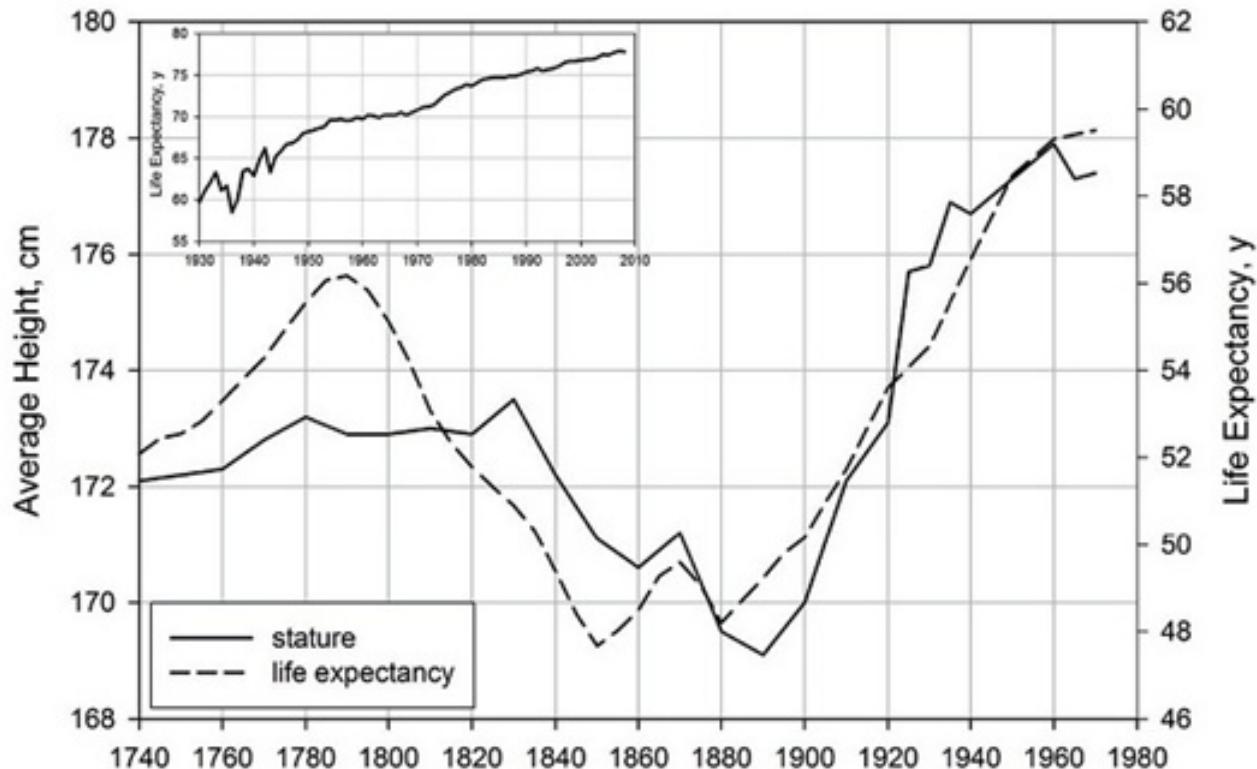
I

I recently reviewed [Secular Cycles](#), which presents a demographic-structural theory of the growth and decline of pre-industrial civilizations. When land is plentiful, population grows and the economy prospers. When land reaches its carrying capacity and income declines to subsistence, the area is at risk of famines, diseases, and wars – which kill enough people that land becomes plentiful again. During good times, elites prosper and act in unity; during bad times, elites turn on each other in an age of backstabbing and civil strife. It seemed pretty reasonable, and authors Peter Turchin and Sergey Nefedov had lots of data to support it. *Ages of Discord* is Turchin's attempt to apply the same theory to modern America. There are many reasons to think this shouldn't work, and the book does a bad job addressing them. So I want to start by presenting Turchin's data showing such cycles exist, so we can at least see why the hypothesis might be tempting. Once we've seen the data, we can decide how turned off we want to be by the theoretical problems.

The first of Turchin's two cyclic patterns is a long cycle of national growth and decline. In *Secular Cycles'* pre-industrial societies, this pattern lasted about 300 years; in *Ages of Discord*'s picture of the modern US, it lasts about 150:



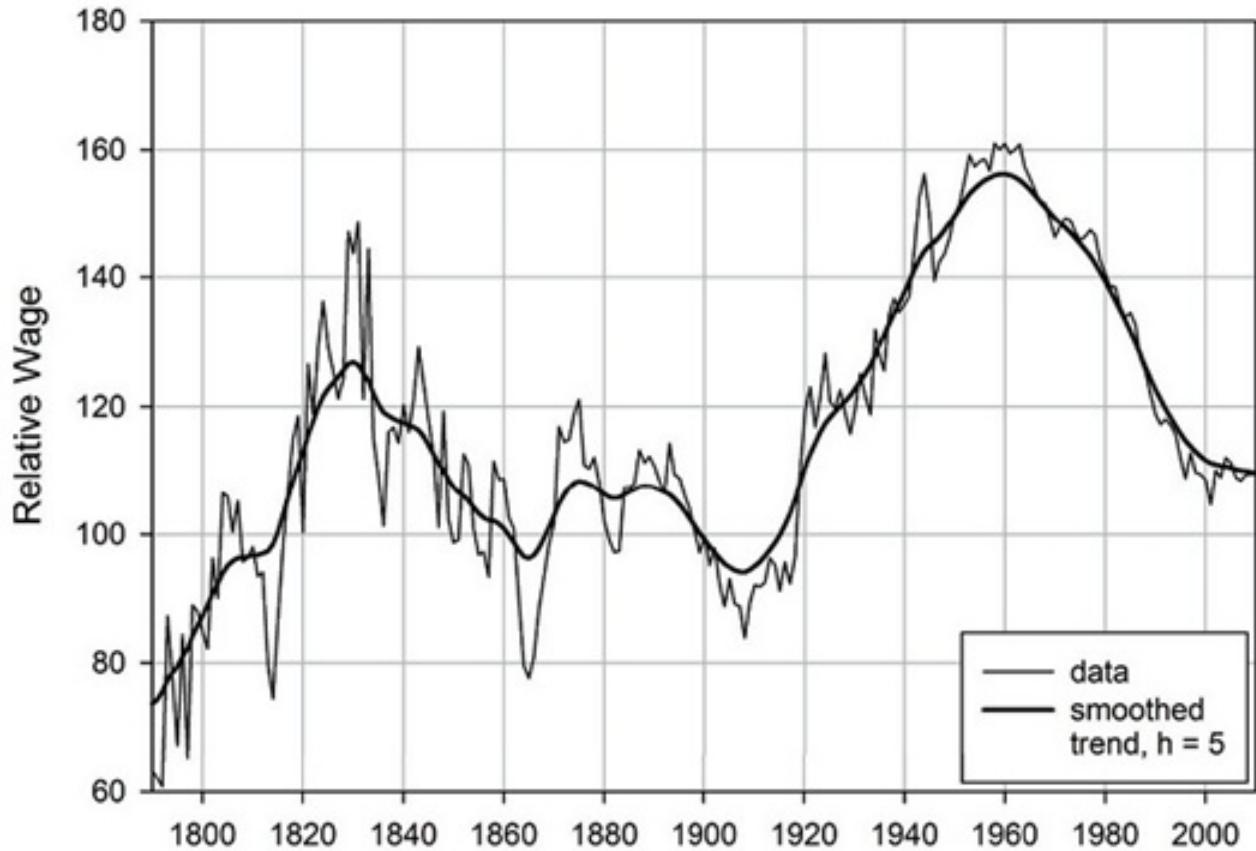
This summary figure combines many more specific datasets. For example, archaeologists frequently assess the prosperity of a period by the heights of its skeletons. Well-nourished, happy children tend to grow taller; a layer with tall skeletons probably represents good times during the relevant archaeological period; one with stunted skeletons probably represents famine and stress. What if we applied this to the modern US?



Average US height and life expectancy over time.

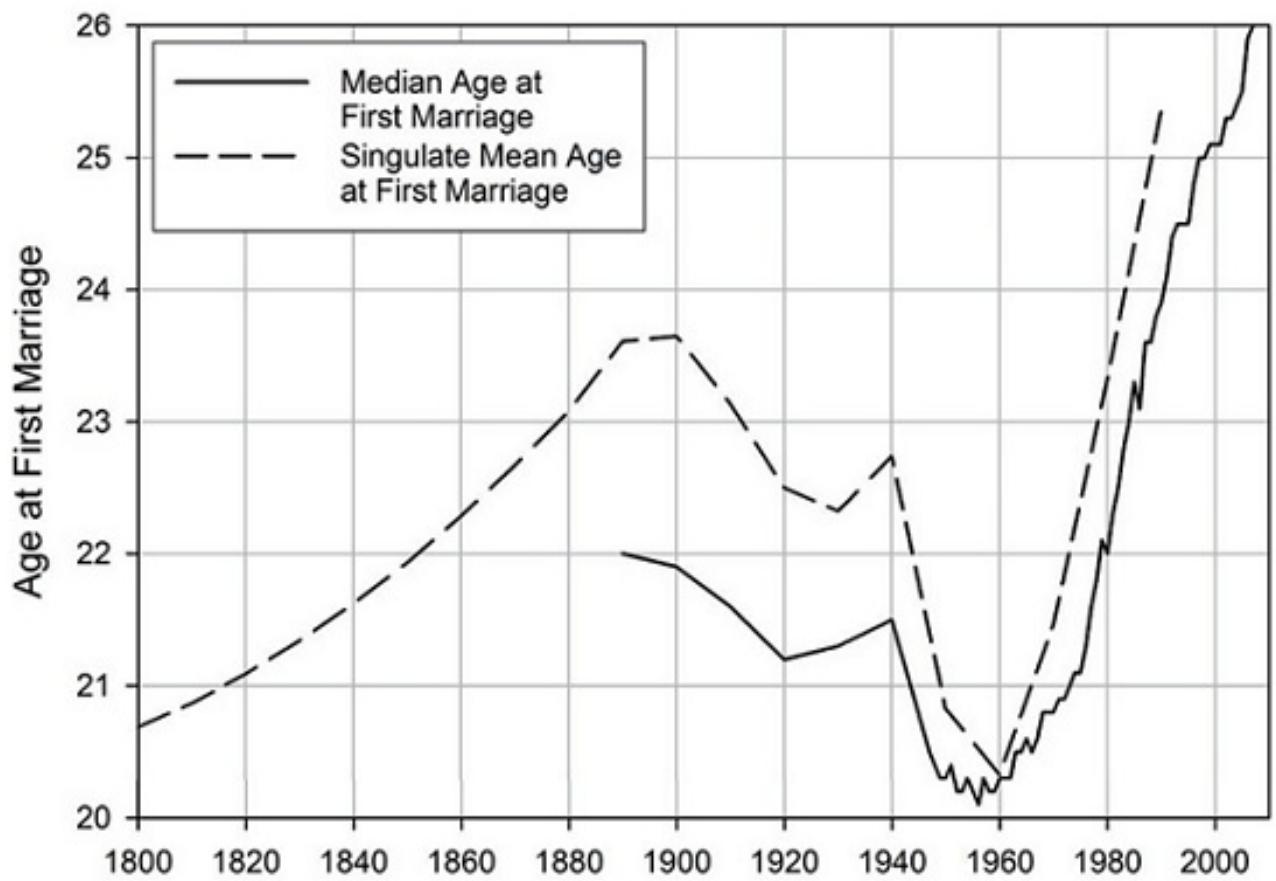
As far as I can tell, the height graph is raw data. The life expectancy graph is the raw data minus an assumed constant positive trend – that is, given that technological advance is increasing life expectancy at a linear rate, what are the other factors you see when you subtract that out? The exact statistical logic is buried in Turchin's source (**Historical Statistics of the United States**, Carter et al 2004), which I don't have and can't judge.

This next graph is the median wage divided by GDP per capita, a crude measure of income equality:

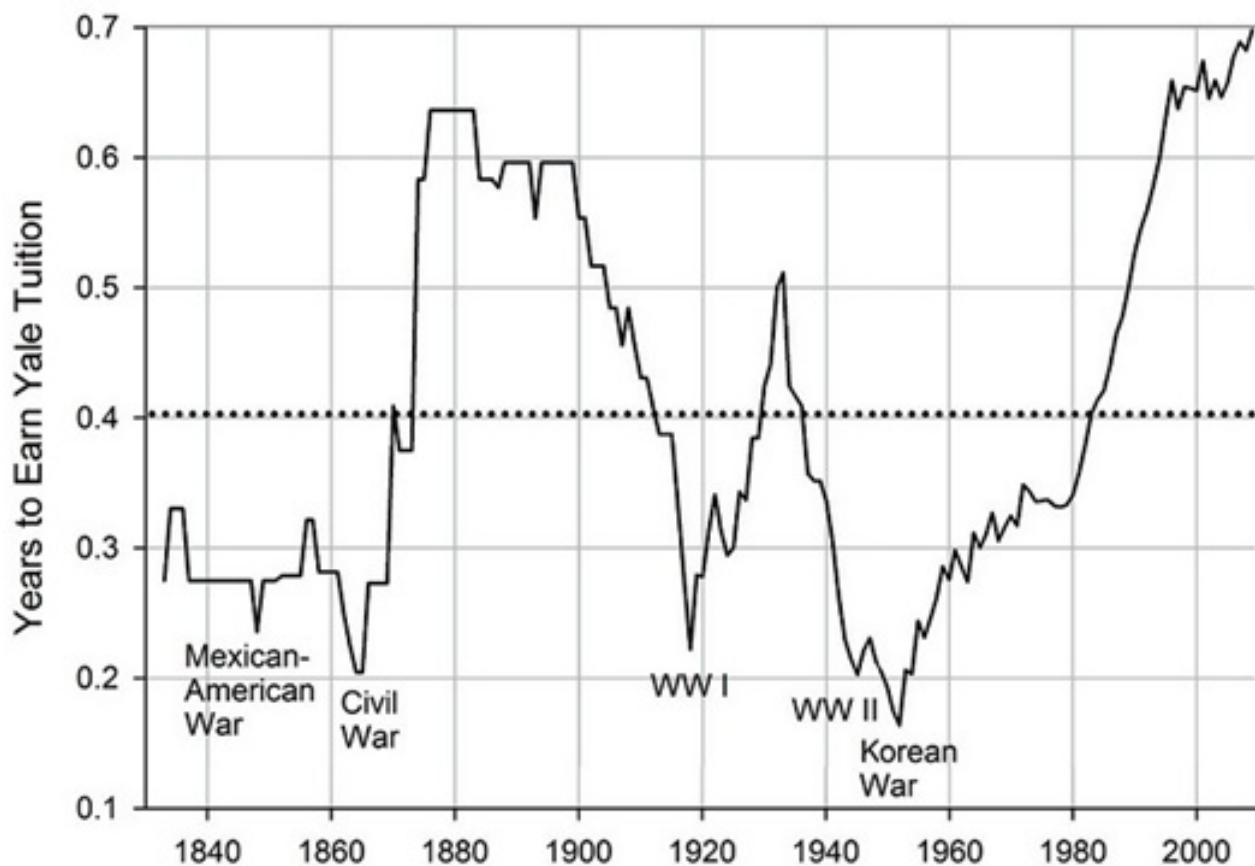


Lower values represent more inequality.

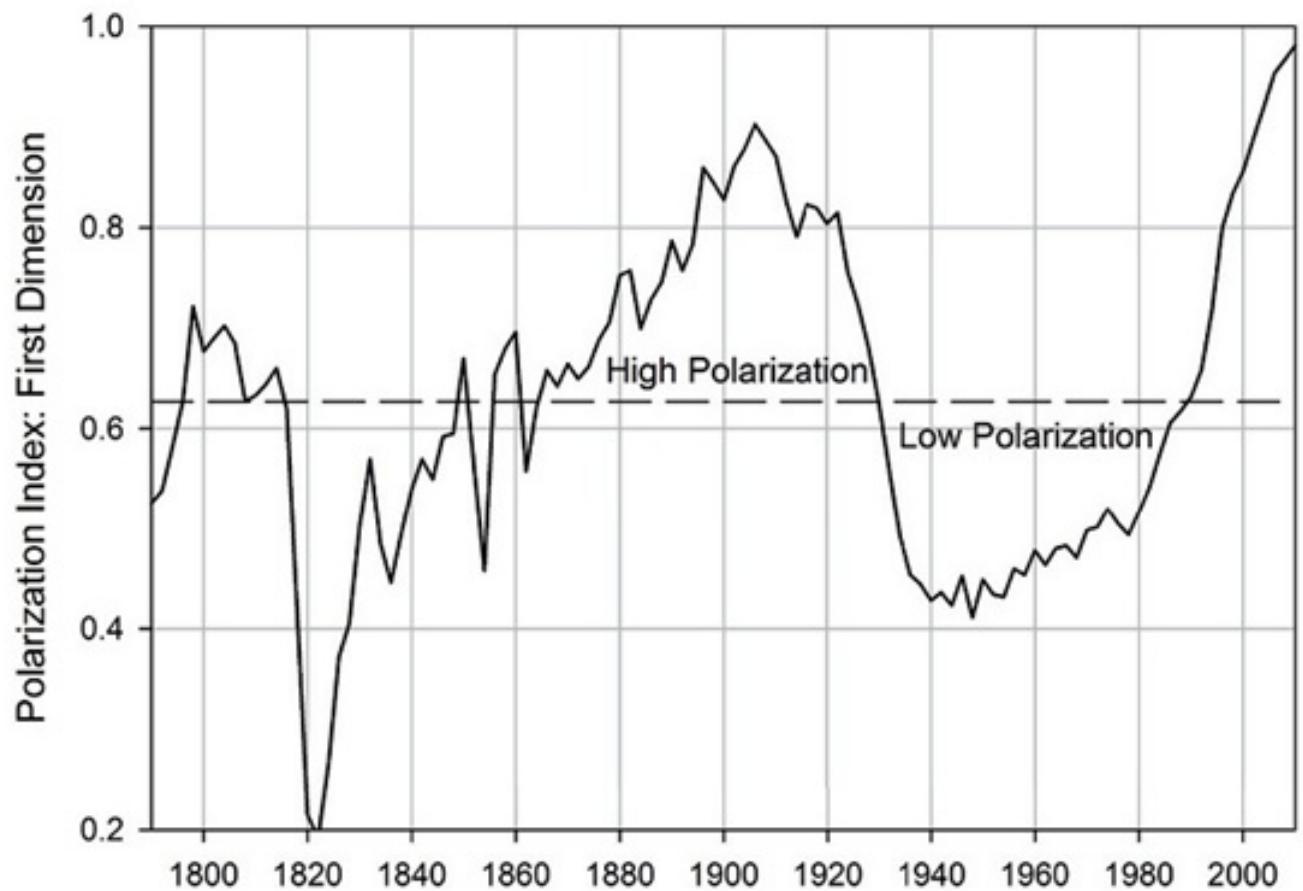
This next graph is median female age at first marriage. Turchin draws on research suggesting this tracks social optimism. In good times, young people can easily become independent and start supporting a family; in bad times, they will want to wait to make sure their lives are stable before settling down:



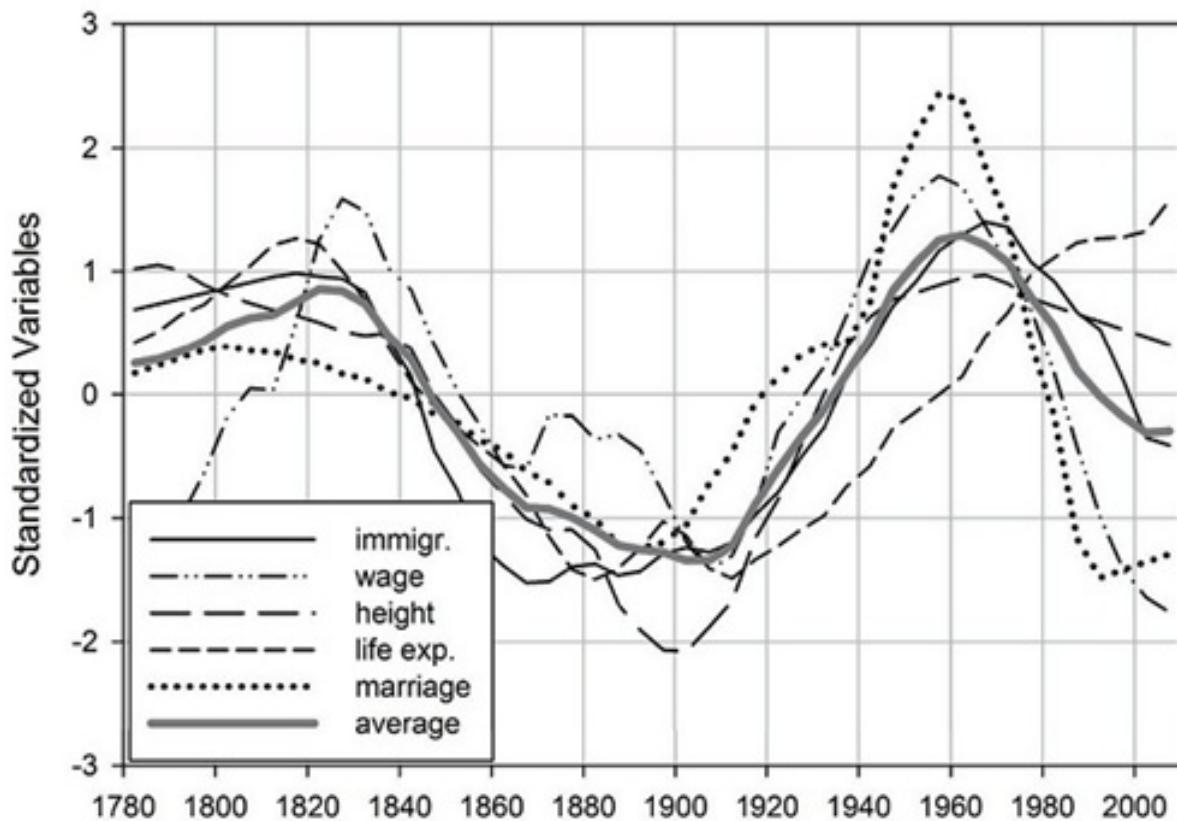
This next graph is Yale tuition as a multiple of average manufacturing worker income. To some degree this will track inequality in general, but Turchin thinks it also measures something like “difficulty of upward mobility”:



This next graph shows DW-NOMINATE's "Political Polarization Index", a complicated metric occasionally used by historians of politics. It measures the difference in voting patterns between the average Democrat in Congress and the average Republican in Congress (or for periods before the Democrats and Republicans, whichever two major parties there were). During times of low partisanship, congressional votes will be dominated by local or individual factors; during times of high partisanship, it will be dominated by party identification:



I've included only those graphs which cover the entire 1780 – present period; the book includes many others that only cover shorter intervals (mostly the more recent periods when we have better data). All of them, including the shorter ones not included here, reflect the same general pattern. You can see it most easily if you standardize all the indicators to the same scale, match the signs so that up always means good and down always means bad, and put them all together:

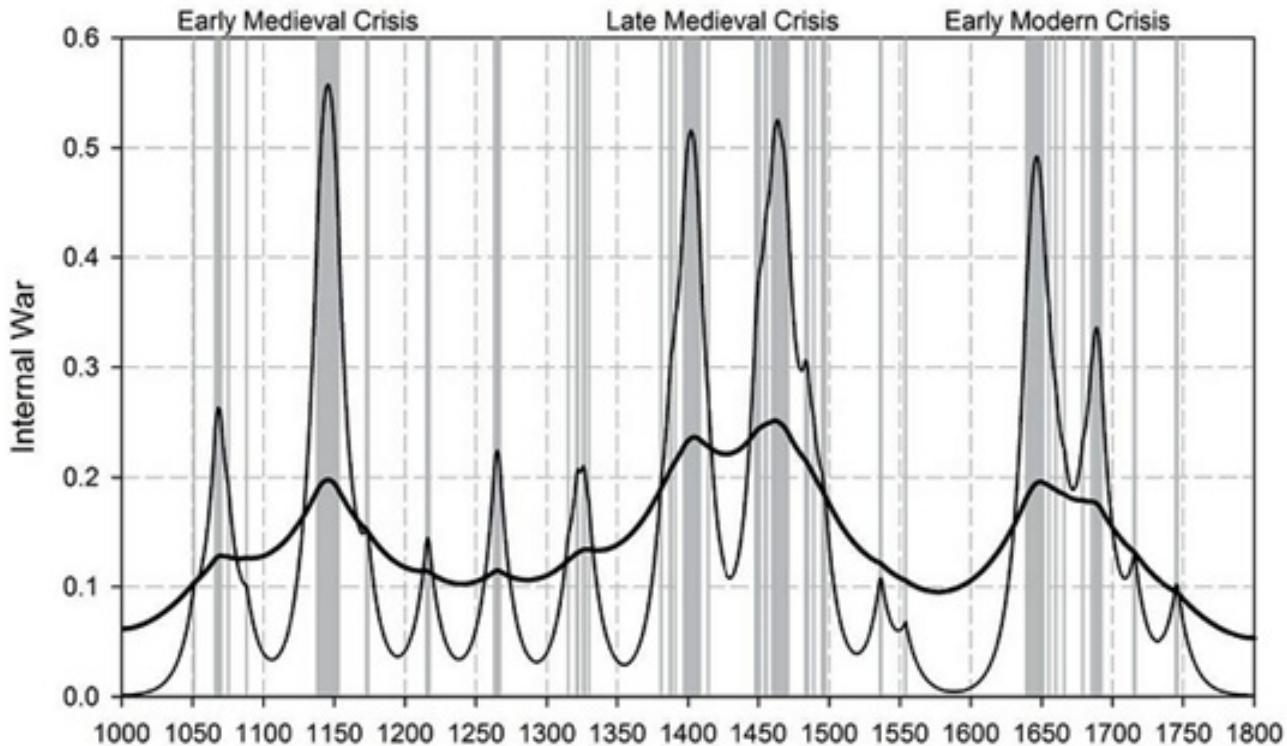


Note that these aren't exactly the same indicators I featured above; we'll discuss immigration later.

The “average” line on this graph is the one that went into making the summary graphic above. Turchin believes that after the American Revolution, there was a period of instability lasting a few decades (eg Shays’ Rebellion, Whiskey Rebellion) but that America reached a maximum of unity, prosperity, and equality around 1820. Things gradually got worse from there, culminating in a peak of inequality, misery, and division around 1900. The reforms of the Progressive Era gradually made things better, with another unity/prosperity/equality maximum around 1960. Since then, an increasing confluence of negative factors (named here as the Reagan Era trend reversal, but Turchin admits it began before Reagan) has been making things worse again.

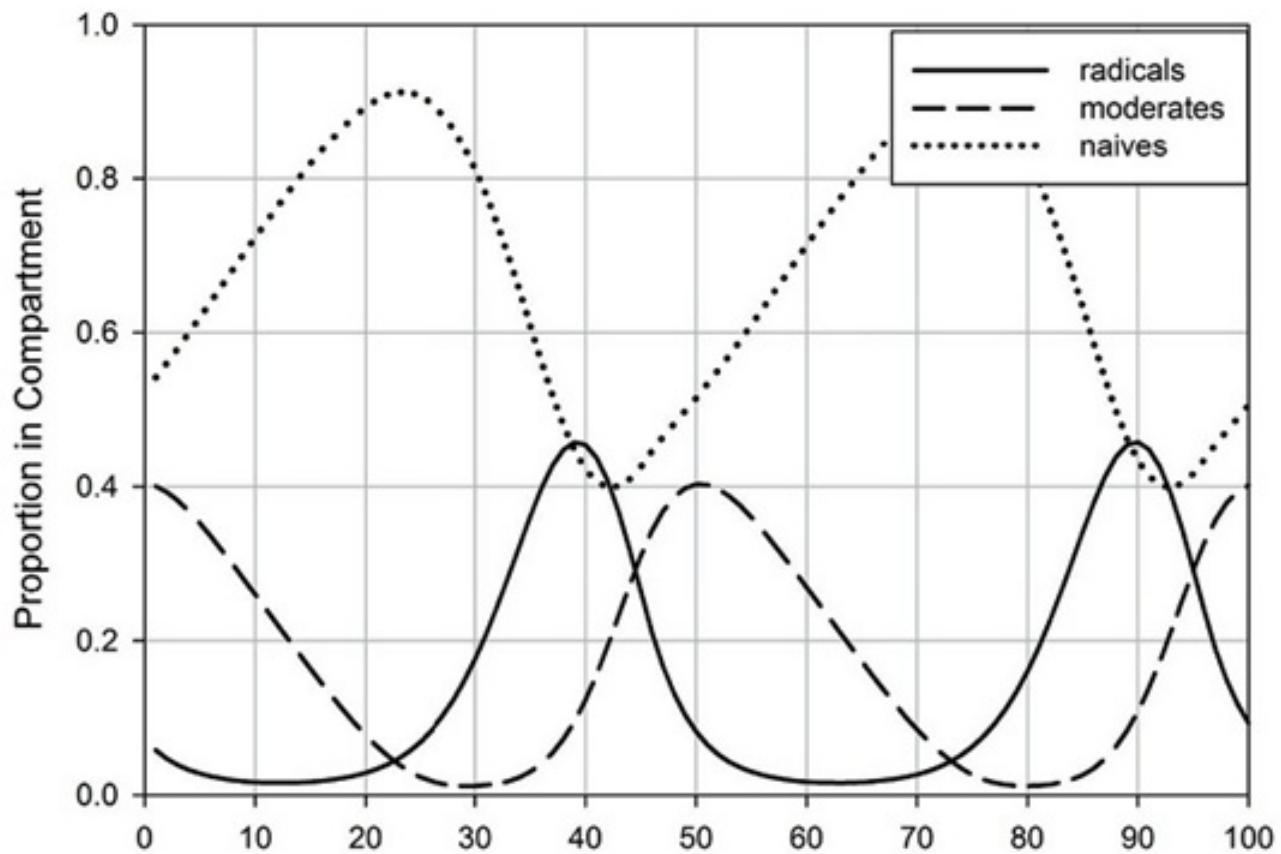
II

Along with this “grand cycle” of 150 years, Turchin adds a shorter instability cycle of 40-60 years. This is the same 40-60 year instability cycle that appeared in *Secular Cycles*, where Turchin called it “the bigenerational cycle”, or the “fathers and sons cycle”.



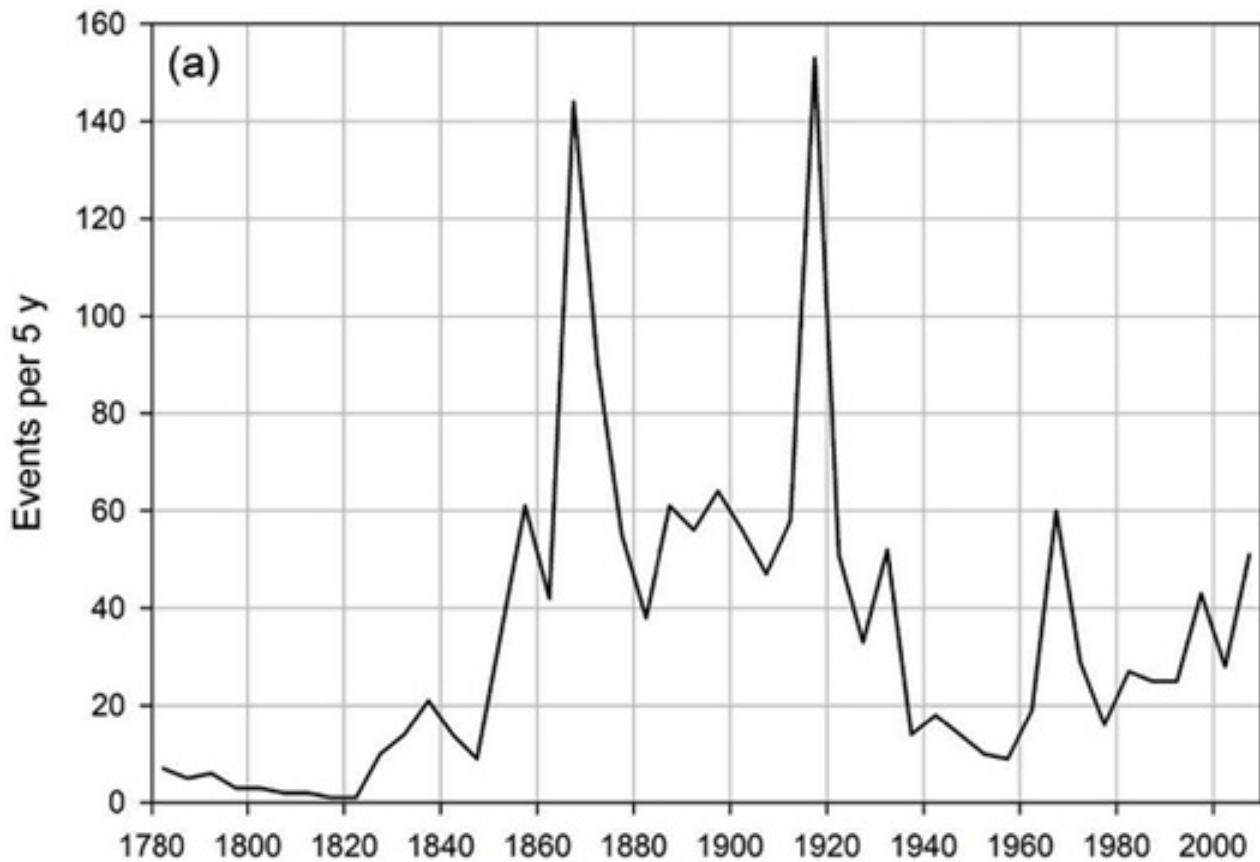
Timing and intensity of internal war in medieval and early modern England, from Turchin and Nefedov 2009.

The derivation of this cycle, explained on pages 45 – 58 of *Ages of Discord*, is one of the highlights of the book. Turchin draws on the kind of models epidemiologists use to track pandemics, thinking of violence as an infection and radicals as plague-bearers. You start with an unexposed vulnerable population. Some radical – patient zero – starts calling for violence. His ideas spread to a certain percent of people he interacts with, gradually “infecting” more and more people with the “radical ideas” virus. But after enough time radicalized, some people “recover” – they become exhausted with or disillusioned by conflict, and become pro-cooperation “active moderates” who are impossible to reinfect (in the epidemic model, they are “inoculated”, but they also have an ability without a clear epidemiological equivalent to dampen radicalism in people around them). As the rates of radicals, active moderates, and unexposed dynamically vary, you get a cyclic pattern. First everyone is unexposed. Then radicalism gradually spreads. Then active moderation gradually spreads, until it reaches a tipping point where it triumphs and radicalism is suppressed to a few isolated reservoirs in the population. Then the active moderates gradually die off, new unexposed people are gradually born, and the cycle starts again. Fiddling with all these various parameters, Turchin is able to get the system to produce 40-60 year waves of instability.



To check this empirically, Turchin tries to measure the number of “instability events” in the US over various periods. He very correctly tries to use lists made by others (since they are harder to bias), but when people haven’t catalogued exactly the kind of instability he’s interested in over the entire 1780 – present period, he sometimes adds his own interpretation. He ends up summing riots, lynchings, terrorism (including assassinations), and mass shootings – you can see his definition for each of these starting on page 114; the short version is that all the definitions seem reasonable but inevitably include a lot of degrees of freedom.

When he adds all this together, here’s what happens:



Political instability / violent events show three peaks, around 1870, 1920, and 1970.

The 1870 peak includes the Civil War, various Civil War associated violence (eg draft riots), and the violence around Reconstruction (including the rise of the Ku Klux Klan and related violence to try to control newly emancipated blacks).

The 1920 peak includes the height of the early US labor movement. Turchin discusses the Mine War, an “undeclared war” from 1920-1921 between bosses and laborers in Appalachian coal country:

Although it started as a labor dispute, it eventually turned into the largest armed insurrection in US history, other than the Civil War. Between 10,000 and 15,000 miners armed with rifles fought thousands of strike-breakers and sheriff's deputies, called the Logan Defenders. The insurrection was ended by the US Army. While such violent incidents were exceptional, they took place against a background of a general “class war” that had been intensifying since the violent teens. “In 1919 nearly four million workers (21% of the workforce) took disruptive action in the face of employer reluctance to recognize or bargain with unions” (Domhoff and Webber, 2011:74).

Along with labor violence, 1920 was also a peak in racial violence:

Race-motivated riots also peaked around 1920. The two most serious such outbreaks were the Red Summer of 1919 (McWhirter 2011) and the Tulsa (Oklahoma) Race Riot. The Red Summer involved riots in more than 20 cities across the United States and resulted in something like 1,000 fatalities. The Tulsa riot in 1921, which caused about 300 deaths, took on an aspect of civil war, in which thousands of whites and

blacks, armed with firearms, fought in the streets, and most of the Greenwood District, a prosperous black neighborhood, was destroyed.

And terrorism:

The bombing campaign by Italian anarchists (“Galleanists”) culminated in the 1920 explosion on Wall Street, which caused 38 fatalities.

The same problems: labor unrest, racial violence, terrorism – repeated during the 1970s spike. Instead of quoting Turchin on this, I want to quote [this Status 451 review](#) of *Days of Rage*, because it blew my mind:

“People have completely forgotten that in 1972 we had over nineteen hundred domestic bombings in the United States.” — Max Noel, FBI (ret.)

Recently, I had my head torn off by a book: Bryan Burrough’s *Days of Rage*, about the 1970s underground. It’s the most important book I’ve read in a year. So I did a series of running tweetstorms about it, and Clark asked me if he could collect them for posterity. I’ve edited them slightly for editorial coherence.

Days of Rage is important, because this stuff is forgotten and it shouldn’t be. The 1970s underground wasn’t small. It was hundreds of people becoming urban guerrillas. Bombing buildings: the Pentagon, the Capitol, courthouses, restaurants, corporations. Robbing banks. Assassinating police. People really thought that revolution was imminent, and thought violence would bring it about.

One thing that Burrough returns to in *Days of Rage*, over and over and over, is how forgotten so much of this stuff is. Puerto Rican separatists bombed NYC like 300 times, killed people, shot up Congress, tried to kill POTUS (Truman). Nobody remembers it.

The passage speaks to me because – yeah, nobody remembers it. This is also how I feel about the 1920 spike in violence. I’d heard about the Tulsa race riot, but the Mine War and the bombing of Wall Street and all the other stuff was new to me. This matters because my intuitions before reading this book would not have been that there were three giant spikes in violence/instability in US history located fifty years apart. I think the lesson I learn is not to trust my intuitions, and to be a little more sympathetic to Turchin’s data.

One more thing: the 1770 spike was obviously the American Revolution and all of the riots and communal violence associated with it (eg against Tories). Where was the 1820 spike? Turchin admits it didn’t happen. He says that because 1820 was the absolute best part of the 150 year grand cycle, everybody was so happy and well-off and patriotic that the scheduled instability peak just fizzled out. Although Turchin doesn’t mention it, you could make a similar argument that the 1870 spike was especially bad (see: the entire frickin’ Civil War) because it hit close to (though not exactly at) the worst part of the grand cycle. 1920 hit around the middle, and 1970 during a somewhat-good period, so they fell in between the non-issue of 1820 and the disaster of 1870.

III

I haven’t forgotten the original question – what drives these 150 year cycles of rise and decline – but I want to stay with the data just a little longer. Again, these data are really interesting. Either some sort of really interesting theory has to be behind them – or they’re just low-quality data cherry-picked to make a point. Which are they? Here are a couple of [spot-checks](#) to see if the data are any good.

First spot check: can I confirm Turchin's data from independent sources?

- [Here](#) is a graph of average US height over time which seems broadly similar to Turchin's.
- [Here](#) is a different measure of US income inequality over time, which again seems broadly similar to Turchin's. [Piketty](#) also presents very similar data, though his story places more emphasis on the World Wars and less on the labor movement.
- The Columbia Law Review measures [political polarization over time](#) and gets mostly the same numbers as Turchin.

I'm going to consider this successfully checked; Turchin's data all seem basically accurate.

Second spot check: do other indicators Turchin didn't include confirm the pattern he detects, or did he just cherry-pick the data series that worked? Spoiler: I wasn't able to do this one. It was too hard to think of measures that should reflect general well-being and that we have 200+ years of unconfounded data for. But here are my various failures:

- The [annual improvement in mortality rate](#) does not seem to follow the cyclic pattern. But isn't this more driven by a few random factors like smoking rates and the logic of technological advance?
- [Treasury bonds](#) maybe kind of follow the pattern until 1980, after which they go crazy.
- [Divorce rates](#) look kind of iffy, but isn't that just a bunch of random factors?
- [Homicide rates](#), with the general downward trend removed, sort of follow the pattern, except for the recent decline?
- [USD/GBP exchange rates](#) don't show the pattern at all, but that could be because of things going on in Britain?

The thing is – really I have no reason to expect divorce rates, homicide rates, exchange rates etc to track national flourishing. For one thing, they may just be totally unrelated. For another, even if they were tenuously related, there are all sorts of other random factors that can affect them. The problem is, I would have said this was true for height, age at first marriage, and income inequality too, before Turchin gave me convincing-sounding stories for why it wasn't. I think my lesson is that I have no idea which indicators should vs. shouldn't follow a secular-cyclic pattern and so I can't do this spot check against cherry-picking the way I hoped.

Third spot check: common sense. Here are some things that stood out to me:

- The Civil War is at a low-ish part of the cycle, but by no means the lowest.
- The Great Depression happened at a medium part of the cycle, when things should have been quickly getting better.
- Even though there was a lot of new optimism with Reagan, continuing through the Clinton years, the cycle does not reflect this at all.

Maybe we can rescue the first and third problem by combining the 150 year cycle with the shorter 50 year cycle. The Civil War was determined by the 50-year cycle having its occasional burst of violence at the same time the 150-year cycle was at a low-ish point. People have good memories of Reagan because the chaos of the 1970 violence burst had ended.

As for the second, Turchin is aware of the problem. He writes:

There is a widely held belief among economists and other social scientists that the 1930s were the “defining moment” in the development of the American politico-economic system (Bordo et al 1998). When we look at the major structural-demographic variables, however, the decade of the 1930s does not seem to be a turning point. Structural-demographic trends that were established during the Progressive Era continues through the 1930s, although some of them accelerated.

Most notably, all the well-being variables that went through trend reversals before the Great Depression – between 1900 and 1920. From roughly 1910 and to 1960 they all increased roughly monotonically, with only one or two minor fluctuations around the upward trend. The dynamics of real wages also do not exhibit a breaking point in the 1930s, although there was a minor acceleration after 1932.

By comparison, he plays up the conveniently-timed (and hitherto unknown to me) depression of the mid-1890s. Quoting Turchin quoting McCormick:

No depression had ever been as deep and tragic as the one that lasted from 1893 to 1897. Millions suffered unemployment, especially during the winters of 1893-4 and 1894-5, and thousands of ‘tramps’ wandered the countryside in search of food [...]

Despite real hardship resulting from massive unemployment, well-being indicators suggest that the human cost of the Great Depression of the 1930s did not match that of the “First Great Depression” of the 1890s (see also Grant 1983:3-11 for a general discussion of the severity of the 1890s depression. Furthermore, while the 1930s are remembered as a period of violent labor unrest, the intensity of class struggle was actually lower than during the 1890s depression. According to the US Political Violence Database (Turchin et al. 2012) there were 32 lethal labor disputes during the 1890s that collectively caused 140 deaths, compared with 20 such disputes in the 1930s with the total of 55 deaths. Furthermore, the last lethal strike in US labor history was in 1937... in other words, the 1930s was actually the last uptick of violent class struggle in the US, superimposed on an overall declining trend.

The 1930s Depression is probably remembered (or rather misremembered) as the worst economic slump in US history, simply because it was the last of the great depressions of the post-Civil War era.

Fourth spot check: Did I randomly notice any egregious errors while reading the book?

On page 70, Turchin discusses “the great cholera epidemic of 1849, which carried away up to 10% of the American population”. This seemed unbelievably high to me. I checked the source he cited, Kohl’s “Encyclopedia Of Plague And Pestilence”, which did give that number. But every other source I checked agreed that the epidemic “only” killed between 0.3% – 1% of the US population (it did hit 10% in a few especially unlucky cities like St. Louis). I cannot fault Turchin’s scholarship in the sense of correctly repeating something written in an encyclopedia, but unless I’m missing something I do fault his common sense.

Also, on page 234, Turchin interprets the percent of medical school graduates who get a residency as “the gap between the demand and supply of MD positions”, which he ties into a wider argument about elite overproduction. But I think this shows a limited understanding of how the medical system works. There is currently a severe *undersupply* of doctors – try getting an appointment with a specialist who takes insurance in a reasonable amount of time if you don’t believe me. Residencies aren’t limited by organic demand. They’re limited because the government places so many restrictions on them that hospitals don’t sponsor them without government funding, and the government is too stingy to fund more of them. None of this has anything to do with elite overproduction.

These are just two small errors in a long book. But they’re two errors in medicine, the field I know something about.

This makes me worry about [Gell-Mann Amnesia](#): if I notice errors in my own field, how many errors must there be in other fields that I just didn't catch?

My overall conclusion from the spot-checks is that the data as presented are basically accurate, but that everything else is so dependent on litigating which things are vs. aren't in accordance with the theory that I basically give up.

IV

Okay. We've gone through the data supporting the grand cycle. We've gone through the data and theory for the 40-60 year instability cycle. We've gone through the reasons to trust vs. distrust the data. Time to go back to the question we started with: why should the grand cycle, originally derived from the Malthusian principles that govern pre-industrial societies, hold in the modern US? Food and land are no longer limiting resources; famines, disease, and wars no longer substantially decrease population. Almost every factor that drives the original secular cycle is missing; why even consider the possibility that it might still apply?

I've put this off because, even though this is the obvious question *Ages of Discord* faces from page one, I found it hard to get a single clear answer.

Sometimes, Turchin talks about the supply vs. demand of labor. In times when the supply of labor outpaces demand, wages go down, inequality increases, elites fragment, and the country gets worse, mimicking the "land is at carrying capacity" stage of the Malthusian cycle. In times when demand for labor exceeds supply, wages go up, inequality decreases, elites unite, and the country gets better. The government is controlled by plutocrats, who always want wages to be low. So they implement policies that increase the supply of labor, especially loose immigration laws. But their actions cause inequality to increase and everyone to become miserable. Ordinary people organize resistance: populist movements, socialist cadres, labor unions. The system teeters on the edge of violence, revolution, and total disintegration. Since the elites don't want those things, they take a step back, realize they're killing the goose that lays the golden egg, and decide to loosen their grip on the neck of the populace. The government becomes moderately pro-labor and progressive for a while, and tightens immigration laws. The oversupply of labor decreases, wages go up, inequality goes down, and everyone is happy. After everyone has been happy for a while, the populists/socialists/unions lose relevance and drift apart. A new generation of elites who have never felt threatened come to power, and they think to themselves "What if we used our control of the government to squeeze labor harder?" Thus the cycle begins again.

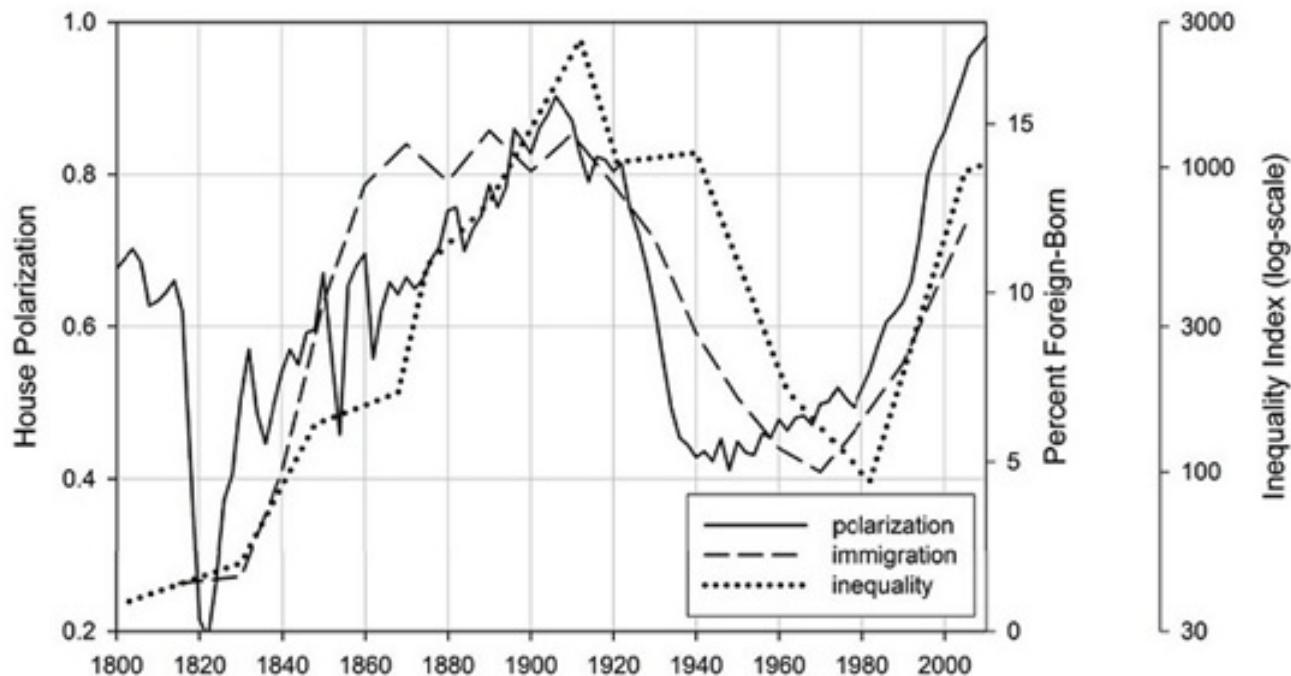
But at other times, Turchin talks more about "elite overproduction". When there are relatively few elites, they can cooperate for their common good. Bipartisanship is high, everyone is unified behind a system perceived as wise and benevolent, and we get a historical period like the 1820s US golden age that historians call [The Era Of Good Feelings](#). But as the number of elites outstrips the number of high-status positions, competition heats up. Elites realize they can get a leg up in an increasingly difficult rat race by backstabbing against each other and the country. Government and culture enter a defect-defect era of hyperpartisanship, where everyone burns the commons of productive norms and institutions in order to get ahead. Eventually... some process reverses this or something?... and then the cycle starts again.

At still other times, Turchin seems to retreat to a sort of mathematical formalism. He constructs an extremely hokey-looking dynamic feedback model, based on ideas like "assume that the level of discontent among ordinary people equals the urbanization rate x the age structure x the inverse of their wages relative to the elite" or "let us define the fiscal distress index as debt ÷ GDP x the level of distrust in state institutions". Then he puts these all together into a model that calculates how the level of discontent affects and is affected by the level of state fiscal distress and a few dozen other variables. On the one hand, this is really cool, and watching it in action gives you the same kind of feeling Seldon must have had inventing psychohistory. On the other, it seems *really* made-up. Turchin admits that

dynamic feedback systems are infamous for going completely haywire if they are even a tiny bit skew to reality, but assures us that he understands the cutting-edge of the field and how to make them not to do that. I don't know enough to judge whether he's right or wrong, but my priors are on "extremely, almost unfathomably wrong". Still, at times he reminds us that the shifts of dynamic feedback systems can be attributed only to the system in its entirety, and that trying to tell stories about or point to specific factors involved in any particular shift is an approximation at best.

All of these three stories run into problems almost immediately.

First, the supply of labor story focuses pretty heavily on immigration. Turchin puts a lot of work into showing that immigration follows the secular cycle patterns; it is highest at the worst part of the cycle, and lowest at the best parts:



In this model, immigration is a tool of the plutocracy. High supply of labor (relative to demand) drives down wages, increases inequality, and lowers workers' bargaining power. If the labor supply is poorly organized, comes from places that don't understand the concept of "union", don't know their rights, and have racial and linguistic barriers preventing them from cooperating with the rest of the working class, well, even better. Thus, periods when the plutocracy is successfully squeezing the working class are marked by high immigration. Periods when the plutocracy fears the working class and feels compelled to be nice to them are marked by low immigration.

This position makes some sense and is loosely supported by the long-term data above. But isn't this one of the most-studied topics in the history of economics? Hasn't it been proven almost beyond doubt that immigrants don't steal jobs from American workers, and that since they consume products themselves (and thus increase the demand for labor) they don't affect the supply/demand balance that sets wages?

It appears I might just be totally miscalibrated on this topic. I checked the [IGM Economic Experts Panel](#). Although most of the expert economists surveyed believed immigration was a net good for America, they did say (50% agree to only 9% disagree) that "unless they were compensated by others, many low-skilled American workers would be substantially worse off if a larger number of low-skilled foreign workers were legally allowed to enter the US each year". I'm having trouble seeing the difference between this statement (which economists seem very convinced is

true) and “you should worry about immigrants stealing your job” (which everyone seems very convinced is false). It might be something like – immigration generally makes “the economy better”, but there’s no guarantee that these gains are evenly distributed, and so it can be bad for low-skilled workers in particular? I don’t know, this would still represent a pretty big update, but given that I was told all top economists think one thing, and now I have a survey of all top economists saying the other, I guess big updates are unavoidable. Interested in hearing from someone who knows more about this.

Even if it’s true that immigration can hurt low-skilled workers, Turchin’s position – which is that increased immigration is responsible for a very large portion of [post-1973 wage stagnation](#) and the recent trend toward rising inequality – sounds shocking to current political sensibilities. But all Turchin has to say is:

An imbalance between labor supply and demand clearly played an important role in driving real wages down after 1978. As Harvard economist George J. Borjas recently wrote, “The best empirical research that tries to examine what has actually happened in the US labor market aligns well with economic theory: An increase in the number of workers leads to lower wages.”

My impression was that Borjas was an increasingly isolated contrarian voice, so once again, I just don’t know what to do here.

Second, the plutocratic oppression story relies pretty heavily on the idea that inequality is a unique bad. This fits the zeitgeist pretty well, but it’s a little confusing. Why should commoners care about their wages relative to elites, as opposed to their absolute wages? Although median-wage-relative-to-GDP has gone down over the past few decades, absolute median wage has gone up – just a little, slowly enough that it’s rightly considered a problem – but it has gone up. Since modern wages are well above 1950s wages, in what sense should modern people feel like they are economically bad off in a way 1950s people didn’t? This isn’t a problem for Turchin’s theory so much as a general mystery, but it’s a general mystery I care about a lot. One answer is that the [cost disease](#) is fueled by a Baumol effect pegged to per capital income (see part 3 [here](#)), and this is a way that increasing elite wealth can absolutely (not relatively) immiserate the lower classes.

Likewise, what about [The Spirit Level Delusion](#) and other resources showing that, across countries, inequality is not particularly correlated with social bads? Does this challenge Turchin’s America-centric findings that everything gets worse along with inequality levels?

Third, the plutocratic oppression story meshes poorly with the elite overproduction story. In elite overproduction, united elites are a sign of good times to come; divided elites means dysfunctional government and potential violence. But as [Pseudoerasmus points out](#), united elites are often united *against* the commoners, and we should expect inequality to be highest at times when the elites are able to work together to fight for a larger share of the pie. But I think this is the opposite of Turchin’s story, where elites unite only to make concessions, and elite unity equals popular prosperity.

Fourth, everything about the elite overproduction story confuses me. Who are “elites”? This category made sense in *Secular Cycles*, which discussed agrarian societies with a distinct titled nobility. But Turchin wants to define US elites in terms of wealth, which follows a continuous distribution. And if you’re defining elites by wealth, it doesn’t make sense to talk about “not enough high-status positions for all elites”; if you’re elite (by virtue of your great wealth), by definition you already have what you need to maintain your elite status. Turchin seems aware of this issue, and sometimes talks about “elite aspirants” – some kind of upper class who expect to be wealthy, but might or might not get that aspiration fulfilled. But then understanding elite overproduction hinges on what makes one non-rich-person person a commoner vs. another non-rich-person an “elite aspirant”, and I don’t remember any clear discussion of this in the book.

Fifth, what drives elite overproduction? Why do elites (as a percent of the population) increase during some periods and decrease during others? Why should this be a cycle rather than a random walk?

My guess is that *Ages of Discord* contains answers to some of these questions and I just missed them. But I missed them after reading the book pretty closely to try to find them, and I didn't feel like there were any similar holes in *Secular Cycles*. As a result, although the book had some fascinating data, I felt like it lacked a clear and lucid thesis about exactly what was going on.

V

Accepting the data as basically right, do we have to try to wring some sense out of the theory?

The data cover a cycle and a half. That means we only sort of barely get to see the cycle "repeat". The conclusion that it is a cycle and not some disconnected trends is based only on the single coincidence that it was 70ish years from the first turning point (1820) to the second (1890), and also 70ish years from the second to the third (1960).

A parsimonious explanation would be "for some reason things were going unusually well around 1820, unusually badly around 1890, and unusually well around 1960 again." This is actually really interesting – I didn't know it was true before reading this book, and it changes my conception of American history a lot. But it's a lot less interesting than the discovery of a secular cycle.

I think the parsimonious explanation is close to what Thomas Piketty argued in his [*Capital In The Twenty-First Century*](#). Inequality was rising until the World Wars, because that's what inequality naturally does given reasonable assumptions about growth rates. Then the Depression and World Wars wiped out a lot of existing money and power structures and made things equal again for a little while. Then inequality started rising again, because that's what inequality naturally does given reasonable assumptions about growth rates. Add in a pinch of *The Spirit Level* – inequality is a mysterious magic poison that somehow makes everything else worse – and there's not much left to be explained.

(some exceptions: why was inequality decreasing until 1820? Does inequality *really* drive political polarization? When immigration corresponds to periods of high inequality, is the immigration *causing* the inequality? And what about the 50 year cycle of violence? That's another coincidence we didn't include in the coincidence list!)

So what can we get from *Ages of Discord* that we can't get from Piketty?

First, the concept of "elite overproduction" is one that worms its way into your head. It's the sort of thing that was constantly in the background of [*Increasingly Competitive College Admissions: Much More Than You Wanted To Know*](#). It's the sort of thing you think about when a million fresh-faced college graduates want to become Journalists and Shape The Conversation and Fight For Justice and realistically just end up getting ground up and spit out by clickbait websites. *Ages of Discord* didn't do a great job breaking down its exact dynamics, but I'm grateful for its work bringing it from a sort of shared unconscious assumption into the light where we can talk about it.

Second, the idea of a deep link between various indicators of goodness and badness – like wages and partisan polarization – is an important one. It forces me to reevaluate things I had considered settled, like that immigration doesn't worsen inequality, or that inequality is not a magical curse that poisons everything.

Third, historians have to choose what events to focus on. Normal historians usually focus on the same normal events. Unusual historians sometimes focus on neglected events that support their unusual theses, so reading someone like Turchin is a good way to learn parts of history you'd never encounter otherwise. Some of these I was able to mention

above – like the Mine War of 1920 or the cholera epidemic of 1849; I might make another post for some of the others.

Fourth, it tries to link events most people would consider separate – wage stagnation since 1973, the Great Stagnation in technology, the decline of Peter Thiel's "[definite optimism](#)", the rise of partisan polarization. I'm not sure exactly how it links them or what it has to say about the link, but link them it does.

But the most important thing about this book is that Turchin claims to be able to predict the future. The book (written just before Trump was elected in 2016) ends by saying that "we live in times of intensifying structural-demographic pressures for instability". The next bigenerational burst of violence is scheduled for about 2020 (realistically +/- a few years). It's at a low point in the grand cycle, so it should be a doozy.

What about beyond that? It's unclear *exactly* where he thinks we are right now in the grand cycle. If the current cycle lasts exactly as long as the last one, we would expect it to bottom out in 2030, but Turchin never claims every cycle is exactly as long. A few of his graphs suggest a hint of curvature, suggesting we might currently be in the worst of it. The socialists seem to have gotten their act together and become an important political force, which the theory predicts is a necessary precursor to change.

I think we can count the book as having made correct predictions if violence spikes in the very near future (are the current number of mass shootings enough to satisfy this requirement? I would have to see it graphed using the same measurements as past spikes), and if sometime in the next decade or so things start looking like there's a ray of light at the end of the tunnel.

I am pretty interested in finding other ways to test Turchin's theories. I'm going to ask some of my math genius friends to see if the dynamic feedback models check out; if anyone wants to help, let me know how I can help you (if money is an issue, I can send you a copy of the book, and I will definitely publish anything you find on this blog). If anyone has any other ideas for indicators that should be correlated with the secular cycle, and ideas about how to find them, I'm interested in that too. And if anyone thinks they can explain the elite overproduction issue, please enlighten me.

I ended my review of *Secular Cycles* by saying:

One thing that strikes me about [Turchin]'s cycles is the ideological component. They describe how, during a growth phase, everyone is optimistic and patriotic, secure in the knowledge that there is enough for everybody. During the stagflation phase, inequality increases, but concern about inequality increases even more, zero-sum thinking predominates, and social trust craters (both because people are actually defecting, and because it's in lots of people's interest to play up the degree to which people are defecting). By the crisis phase, partisanship is much stronger than patriotism and radicals are talking openly about how violence is ethically obligatory.

And then, eventually, things get better. There is a new Augustan Age of virtue and the reestablishment of all good things. This is a really interesting claim. Western philosophy tends to think in terms of trends, not cycles. We see everything going on around us, and we think this is some endless trend towards more partisanship, more inequality, more hatred, and more state dysfunction. But *Secular Cycles* offers a narrative where endless trends can end, and things can get better after all.

This is still the hope, I guess. I don't have a lot of faith in human effort to restore [niceness, community, and civilization](#). All I can do is pray the Vast Formless Things accomplish it for us without asking us first.

1960: The Year The Singularity Was Cancelled

Posted on April 22, 2019 by Scott Alexander



Epistemic status: Very speculative, especially Parts 3 and 4. Like many good things, this post is based on a conversation with Paul Christiano; most of the good ideas are his, any errors are mine.

I

In the 1950s, an Austrian scientist discovered a series of equations that he claimed could model history. They matched past data with startling accuracy. But when extended into the future, they predicted the world would end on November 13, 2026.

This sounds like the plot of a sci-fi book. But it's also the story of [Heinz von Foerster](#), a mid-century physicist, cybernetician, cognitive scientist, and philosopher.

His problems started when he became interested in human population dynamics.

(the rest of this section is loosely adapted from his *Science* paper "[Doomsday: Friday, 13 November, A.D. 2026](#)")

Assume a perfect paradisiacal Garden of Eden with infinite resources. Start with two people – Adam and Eve – and assume the population doubles every generation. In the second generation there are 4 people; in the third, 8. This is that old riddle about the [grains of rice on the chessboard](#) again. By the 64th generation (ie after about 1500 years) there will be 18,446,744,073,709,551,616 people – ie about about a billion times the number of people who have ever lived in all the eons of human history. So one of our assumptions must be wrong. Probably it's the one about the perfect paradise with unlimited resources.

Okay, new plan. Assume a world with a limited food supply / limited carrying capacity. If you want, imagine it as an island where everyone eats coconuts. But there are only enough coconuts to support 100 people. If the population reproduces beyond 100 people, some of them will starve, until they're back at 100 people. In the second generation, there are 100 people. In the third generation, still 100 people. And so on to infinity. Here the population never grows at all. But that doesn't match real life either.

But von Foerster knew that technological advance can change the carrying capacity of an area of land. If our hypothetical islanders discover new coconut-tree-farming techniques, they may be able to get twice as much food, increasing the maximum population to 200. If they learn to fish, they might open up entirely new realms of food production, increasing population into the thousands.

So the rate of population growth is neither the double-per-generation of a perfect paradise, nor the zero-per-generation of a stagnant island. Rather, it depends on the rate of economic and technological growth. In particular, in a closed system that is already at its carrying capacity and with zero marginal return to extra labor, population growth equals productivity growth.

What causes productivity growth? Technological advance. What causes technological advance? Lots of things, but von

Foerster's model reduced it to one: people. Each person has a certain percent chance of coming up with a new discovery that improves the economy, so productivity growth will be a function of population.

So in the model, the first generation will come up with some small number of technological advances. This allows them to spawn a slightly bigger second generation. This new slightly larger population will generate slightly more technological advances. So each generation, the population will grow at a slightly faster rate than the generation before.

This matches reality. The world population barely increased at all in the millennium from 2000 BC to 1000 BC. But it doubled in the fifty years from 1910 to 1960. In fact, using his model, von Foerster was able to come up with an equation that predicted the population near-perfectly from the Stone Age until his own day.

But his equations corresponded to something called hyperbolic growth. In hyperbolic growth, a feedback cycle – in this case population causes technology causes more population causes more technology – leads to growth increasing rapidly and finally shooting to infinity. Imagine a simplified version of Foerster's system where the world starts with 100 million people in 1 AD and a doubling time of 1000 years, and the doubling time decreases by half after each doubling. It might predict something like this:

Year	Population
1 AD	100 million people
1000 AD	200 million people
1500 AD	400 million people
1750 AD	800 million people
1875 AD	1600 million people

...and so on. This system reaches infinite population in finite time (ie before the year 2000). The real model that von Foerster got after analyzing real population growth was pretty similar to this, except that it reached infinite population in 2026, give or take a few years (his pinpointing of Friday November 13 was mostly a joke; the equations were not really that precise).

What went wrong? Two things.

First, as von Foerster knew (again, it was kind of a joke) the technological advance model isn't literally true. His hyperbolic model just operates as an upper bound on the Garden of Eden scenario. Even in the Garden of Eden, population can't do more than double every generation.

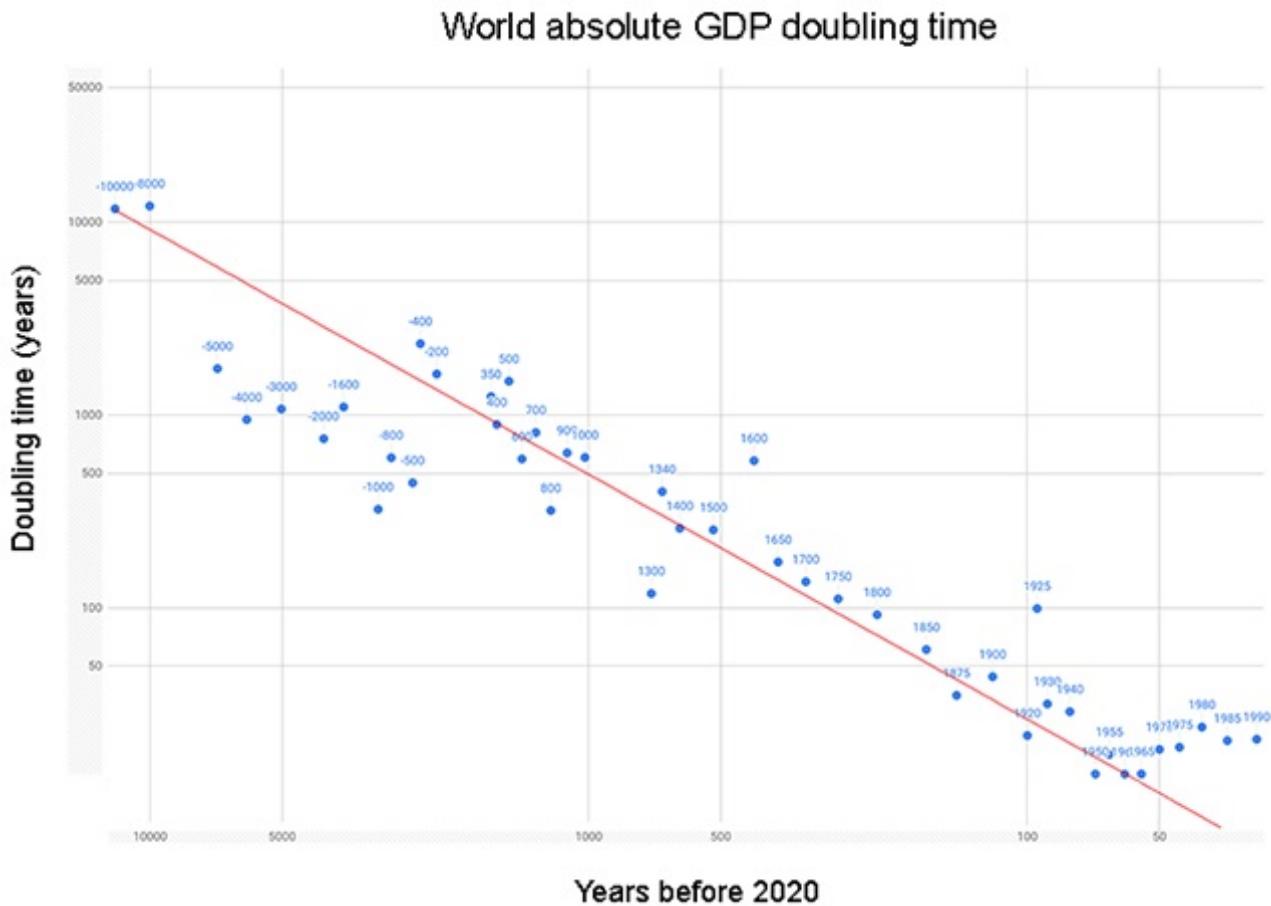
Second, contra all previous history, people in the 1900s started to have fewer kids than their resources could support ([the demographic transition](#)). Couples started considering the cost of college, and the difficulty of maternity leave, and all that, and decided that maybe they should stop at 2.5 kids (or just get a puppy instead).

Von Foerster published his paper in 1960, which ironically was the last year that his equations held true. Starting in 1961, population left its hyperbolic growth path. It is now expected to stabilize by the end of the 21st century.

II

But nobody really expected the population to reach infinity. Armed with this story, let's look at something more interesting.

This ([source](#)) might be the most depressing graph ever:



The horizontal axis is years before 2020, a random year chosen so that we can put this in log scale without negative values screwing everything up. This is an arbitrary choice, but you can also graph it [with log GDP as the horizontal axis](#) and find a similar pattern.

The vertical axis is the amount of time it took the world economy to double from that year, according to [this paper](#). So for example, if at some point the economy doubled every twenty years, the dot for that point is at twenty. The doubling time decreases throughout most of the period being examined, indicating hyperbolic growth.

Hyperbolic growth, as mentioned before, shoots to infinity at some specific point. On this graph, that point is represented by the doubling time reaching zero. Once the economy doubles every zero years, you might as well call it infinite.

For all of human history, economic progress formed a near-perfect straight line pointed at the early 21st century. Its destination varied by a century or two now and then, but never more than that. If an ancient Egyptian economist had modern techniques and methodologies, he could have made a graph like this and predicted it would reach infinity around the early 21st century. If a Roman had done the same thing, using the economic data available in his own time, he would have predicted the early 21st century too. A medieval Burugundian? Early 21st century. A Victorian Englishman? Early 21st century. A Stalinist Russian? Early 21st century. The trend was *really* resilient.

In 2005, inventor Ray Kurzweil published *The Singularity Is Near*, claiming there would be a technological singularity in the early 21st century. He didn't refer to this graph specifically, but he highlighted this same trend of everything

getting faster, including rates of change. Kurzweil took the infinity at the end of this graph very seriously; he thought that some event would happen that really *would* catapult the economy to infinity. Why not? Every data point from the Stone Age to the Atomic Age agreed on this.

This graph shows the Singularity getting cancelled.

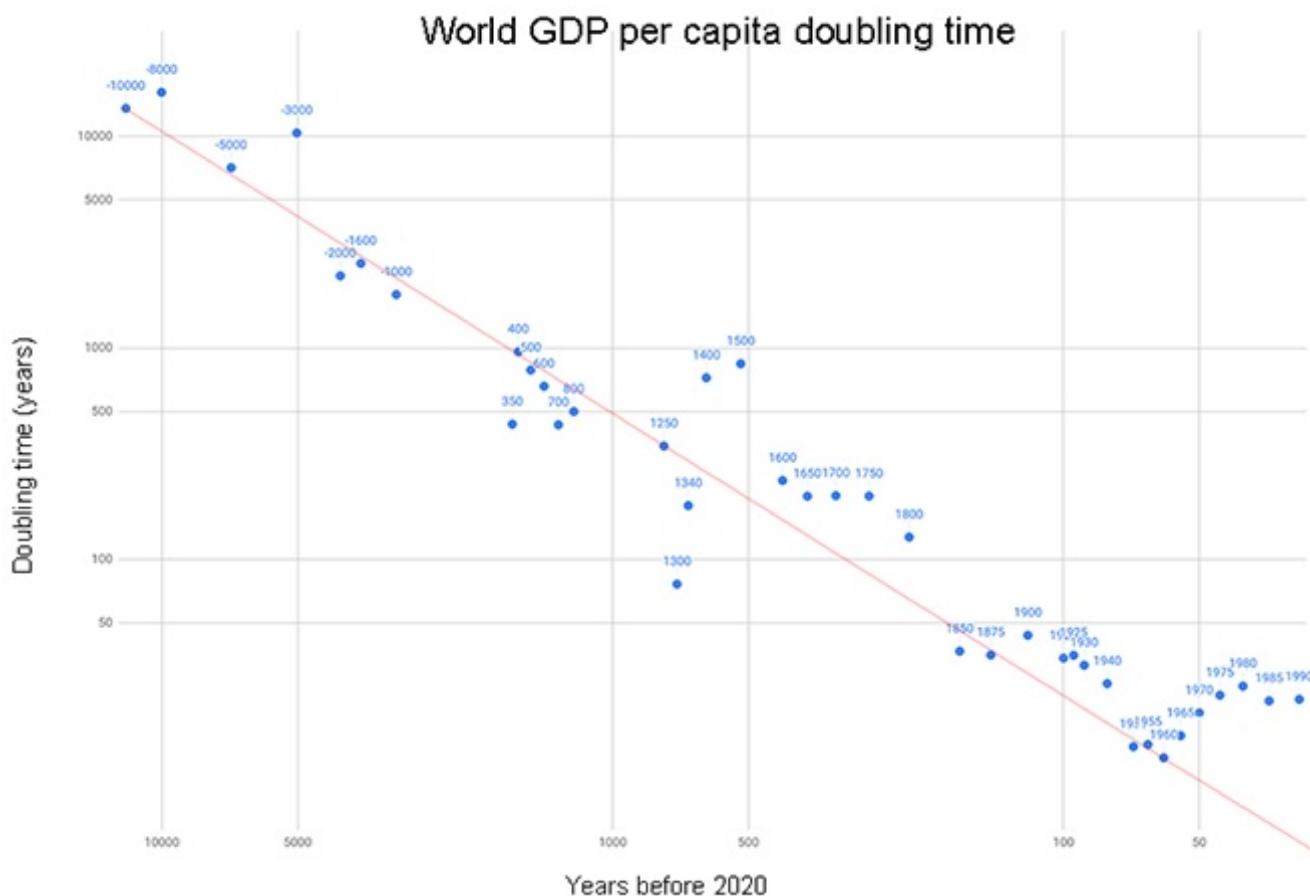
Around 1960, doubling times stopped decreasing. The economy kept growing. But now it grows at a flat rate. It shows no signs of reaching infinity; not soon, not ever. Just constant, boring 2% GDP growth for the rest of time.

Why?

Here von Foerster has a ready answer prepared for us: population!

Economic growth is a function of population and productivity. And productivity depends on technological advancement and technological advancement depends on population, so it all bottoms out in population in the end. And population looked like it was going to grow hyperbolically until 1960, after which it stopped. That's why hyperbolic economic growth, ie progress towards an economic singularity, stopped then too.

In fact...



This is a *really sketchy* graph of per capita income doubling times. It's sketchy because until 1650, per capita income wasn't really increasing at all. It was following a one-step-forward one-step-back pattern. But if you take out all the steps back and just watch how quickly it took the steps forward, you get something like this.

Even though per capita income tries to abstract out population, it displays the same pattern. Until 1960, we were on track for a singularity where everyone earned infinite money. After 1960, the graph “bounces back” and growth rates stabilize or even decrease.

Again, von Foerster can explain this to us. Per capita income grows when technology grows, and technology grows when the population grows. The signal from the end of hyperbolic population growth shows up here too.

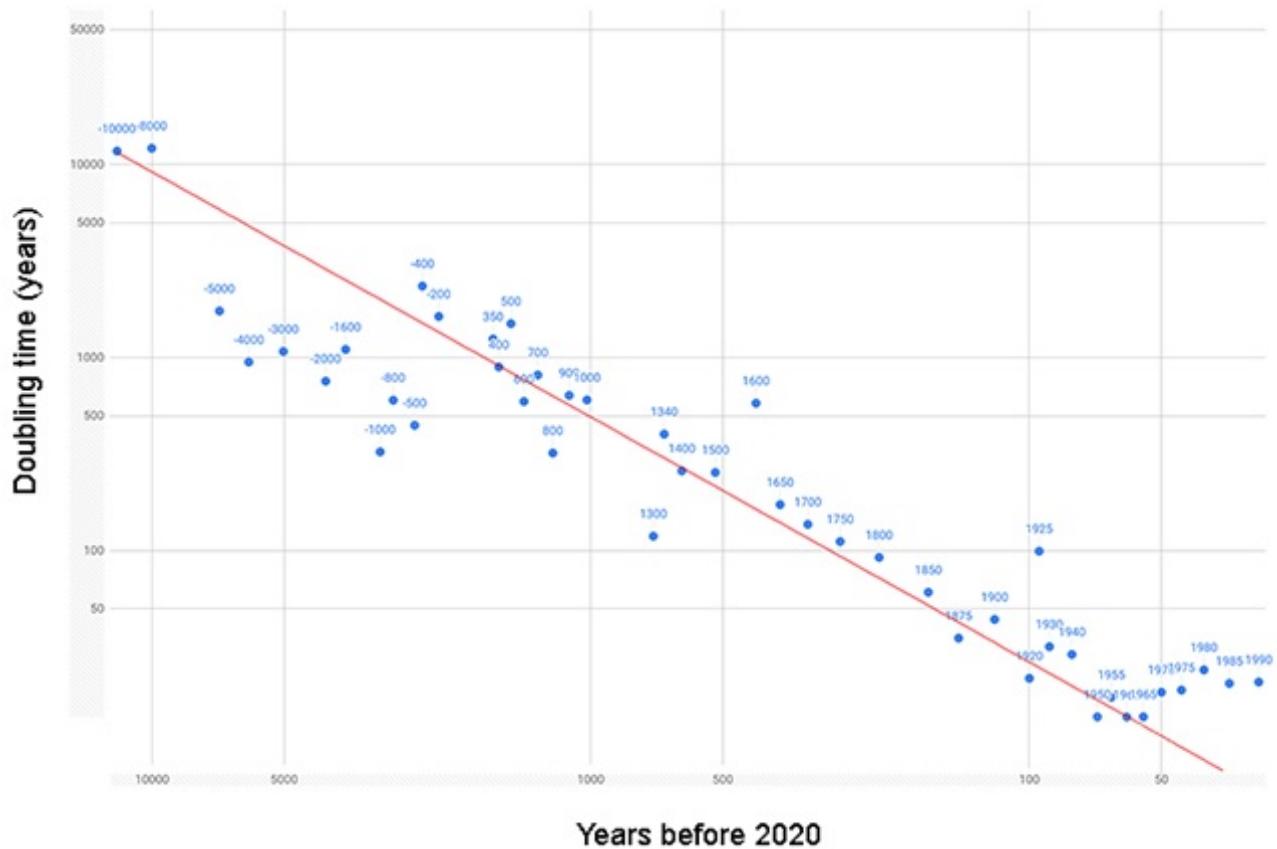
To make this really work, we probably have to zoom in a little bit and look at concrete reality. Most technological advances come from a few advanced countries whose population stabilized a little earlier than the world population. Of the constant population, [an increasing fraction are becoming researchers each year](#) (on the other hand, the low-hanging fruit gets picked off and technological advance becomes harder with time). All of these factors mean we shouldn't expect productivity growth/GWP per capita growth/technological growth to *exactly* track population growth. But on the sort of orders-of-magnitude scale you can see on logarithmic graphs like the ones above, it should be pretty close.

So it looks like past predictions of a techno-economic singularity for the early 21st century were based on extrapolations of a hyperbolic trend in technology/economy that depended on a hyperbolic trend in population. Since the population singularity didn't pan out, we shouldn't expect the techno-economic singularity to pan out either. In fact, since population in advanced countries is starting to “stagnate” relative to earlier eras, we should expect a relative techno-economic stagnation too.

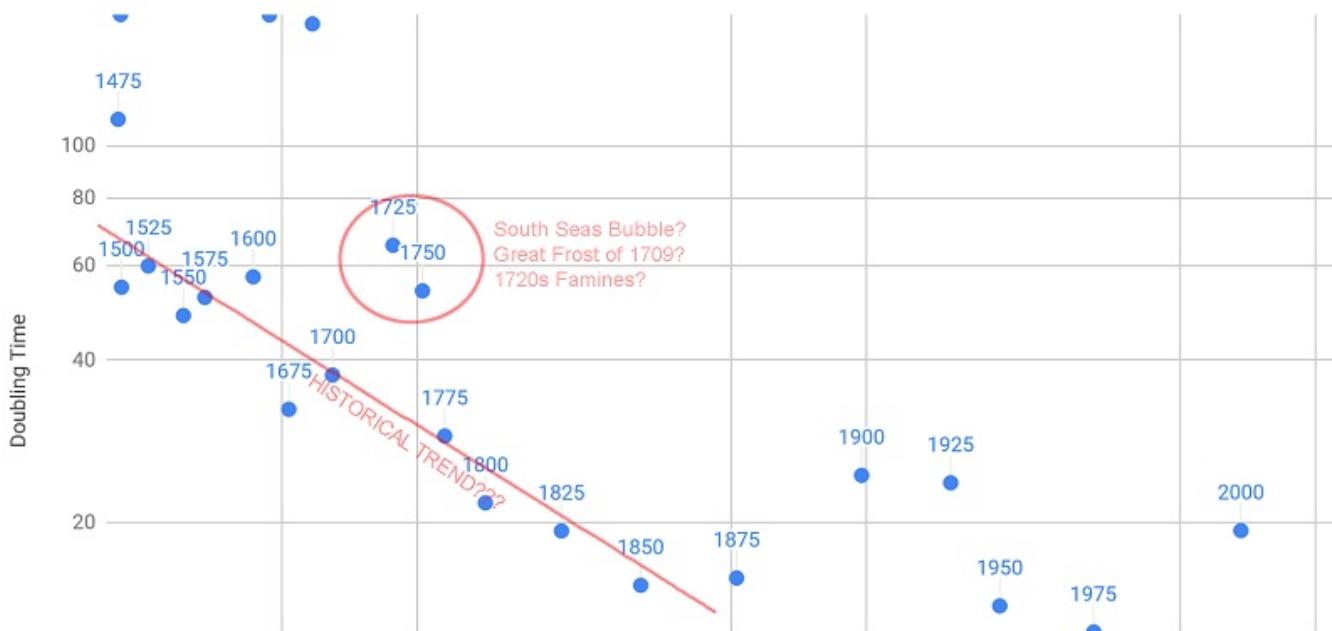
...maybe. Before coming back to this, let's explore some of the other implications of these models.

III

World absolute GDP doubling time



Doubling Time vs. GDP (Britain)



The first graph is the same one you saw in the last section, of absolute GWP doubling times. The second graph is the same, but limited to Britain.

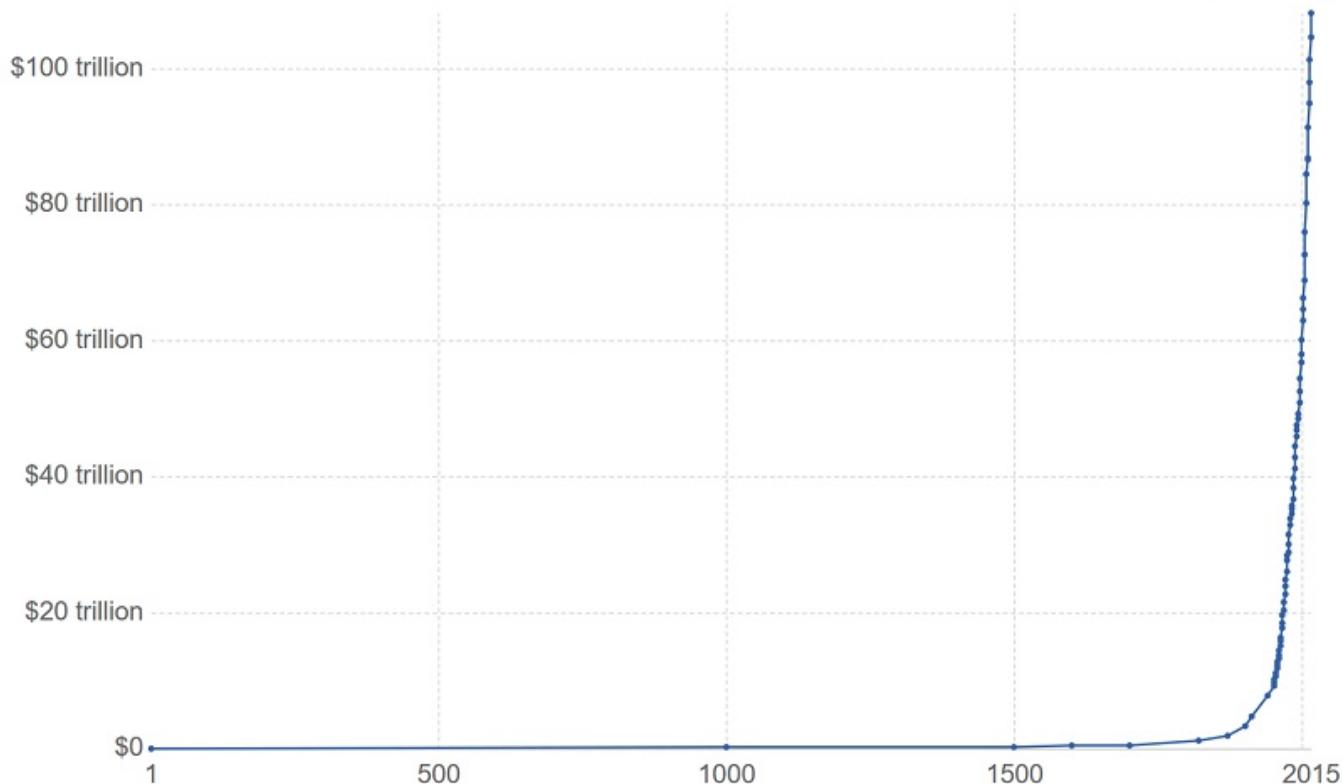
Where's the Industrial Revolution?

It doesn't show up at all. This may be a surprise if you're used to the standard narrative where the Industrial Revolution was the most important event in economic history. Graphs like this make the case that the Industrial Revolution was an explosive shift to a totally new growth regime:

World GDP over the last two millennia

Total output of the world economy; adjusted for inflation and expressed in 2011 international dollars.

OurWorld
in Data



It sure *looks* like the Industrial Revolution was a big deal. But Paul Christiano argues your eyes may be deceiving you. That graph is a hyperbola, ie corresponds to a single simple equation. There is no break in the pattern at any point. If you transformed it to a log doubling time graph, you'd just get the graph above that looks like a straight line until 1960.

On this view, the Industrial Revolution didn't change historical GDP trends. It just shifted the world from a Malthusian regime where economic growth increased the population to a modern regime where economic growth increased per capita income.

For the entire history of the world until 1000, GDP per capita was the same for everyone everywhere during all historical eras. An Israelite shepherd would have had about as much stuff as a Roman farmer or a medieval serf.

This was the Malthusian trap, where "productivity produces people, not prosperity". People reproduce to fill the resources available to them. Everyone always lives at subsistence level. If productivity increases, people reproduce, and now you have more people living at subsistence level. [OurWorldInData has](#) an awesome graph of this:

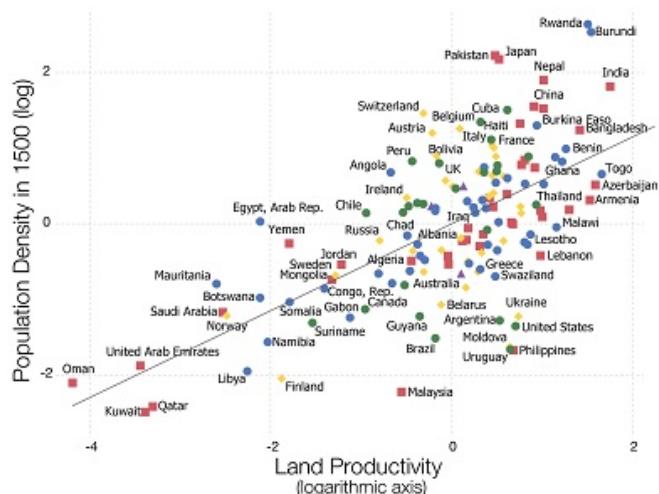
In the Malthusian Economy productivity produces people not prosperity

This figure depicts the partial regression line for the effect of land productivity on income per capita in the year 1500 CE, while controlling for the influence of land productivity, absolute latitude, access to waterways, and continental fixed effects.

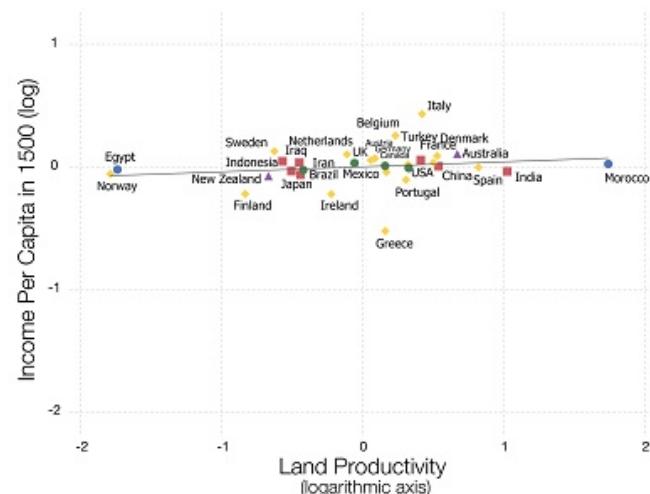
The x- and y-axes plot the residuals obtained from regressing land productivity and income per capita, respectively, on the aforementioned set of covariates.

The color represents the continent of the country: ● Africa ● Europe ■ Asia ▲ Oceania ● Americas

The partial effect of land productivity on population density in 1500 CE



The partial effect of land productivity on income per capita in 1500 CE



Data source: Quamrul Ashraf and Oded Galor (2011) – Dynamics and Stagnation in the Malthusian Epoch. American Economic Review, 101(5): 2003-41.

This is a data visualization from OurWorldInData.org. There you find more visualizations and research on how the world is changing.

Licensed under CC-BY-SA by the author Max Roser.

As of 1500, places with higher productivity (usually richer farmland, but better technology and social organization also help) population density is higher. But GDP per capita was about the same everywhere.

There were always occasional windfalls from exciting discoveries or economic reforms. For a century or two, GDP per capita would rise. But population would always catch up again, and everyone would end up back at subsistence.

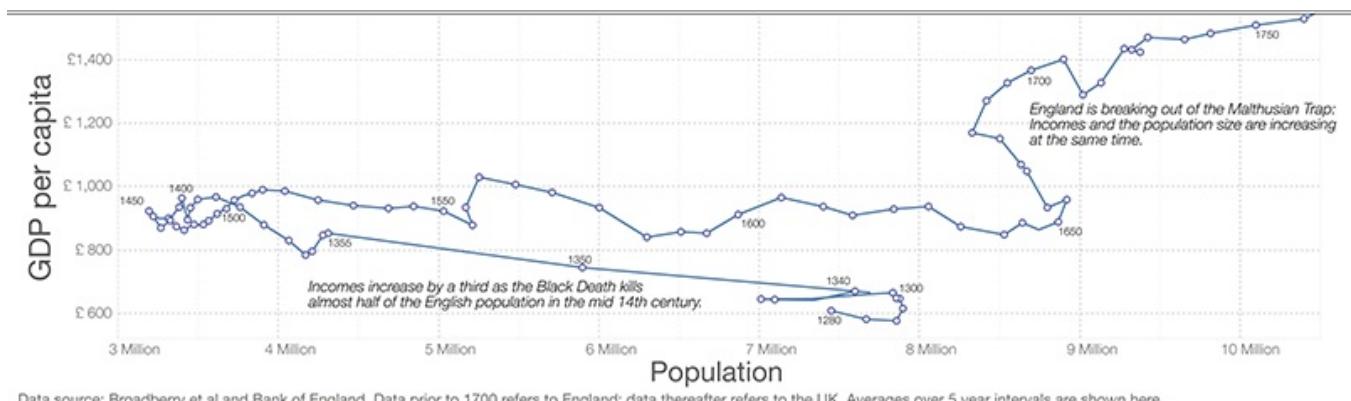
Some people argue Europe broke out of the Malthusian trap around 1300. This is not quite right. 1300s Europe achieved above-subsistence GDP, but only because the Black Plague killed so many people that the survivors got a windfall by taking their land.

Malthus predicts that this should only last a little while, until the European population bounces back to pre-Plague levels. This prediction was exactly right for Southern Europe. Northern Europe didn't bounce back. Why not?

Unclear, but one answer is: fewer people, more plagues. [Broadberry 2015](#) mentions that Northern European culture promoted later marriage and fewer children:

The North Sea Area had an advantage in this area because of its approach to marriage. Hajnal (1965) argued that northwest Europe had a different demographic regime from the rest of the world, characterised by later marriage and hence limited fertility. Although he originally called this the European Marriage Pattern, later work established that it applied [only to the northwest](#) of the continent. This can be linked to the availability of labour market opportunities for females, who could engage in market activity before marriage, thus increasing the age of first marriage for females and reducing the number of children conceived (de Moor and van Zanden, 2010). Later marriage and fewer children are associated with more investment in human capital, since the women employed in productive work can accumulate skills, and parents can afford to invest more in each of the smaller number of children because of the “quantity-quality” trade-off (Voigtländer and Voth, 2010).

This low birth rate was happening at the same time plagues were raising the death rate. Here's another amazing graph from OurWorldInData:



Data source: Broadberry et al and Bank of England. Data prior to 1700 refers to England; data thereafter refers to the UK. Averages over 5 year intervals are shown here.

The visualization is available at OurWorldInData.org where you find more visualizations and research on global development.

Licensed under CC-BY-SA by the author Max Roser.

British population maxes out around 1300 (?), declines substantially during the Black Plague of 1348-49, but then keeps declining. The [List Of English Plagues](#) says another plague hit in 1361, then another in 1369, then another in 1375, and so on. Some historians call the whole period from 1348 to 1666 “the Plague Years”.

It looks like through the 1350 – 1450 period, population keeps declining, and per capita income keeps going up, as Malthusian theory would predict.

Between 1450 and 1550, population starts to recover, and per capita incomes start going down, again as Malthus would predict. Then around 1560, there's a jump in incomes; according to the List Of Plagues, 1563 was “probably the worst of the great metropolitan epidemics, and then extended as a major national outbreak”. After 1563, population increases again and per capita incomes decline again, all the way until 1650. Population does not increase in Britain at all between 1660 and 1700. Why? The List declares 1665 to be “The Great Plague”, the largest in England since 1348.

So from 1348 to 1650, Northern European per capita incomes diverged from the rest of the world's. But they didn't “break out of the Malthusian trap” in a strict sense of being able to direct production toward prosperity rather than population growth. They just had so many plagues that they couldn't grow the population anyway.

But in 1650, England did start breaking out of the Malthusian trap; population and per capita incomes grow together. Why?

Paul theorizes that technological advance finally started moving faster than maximal population growth.

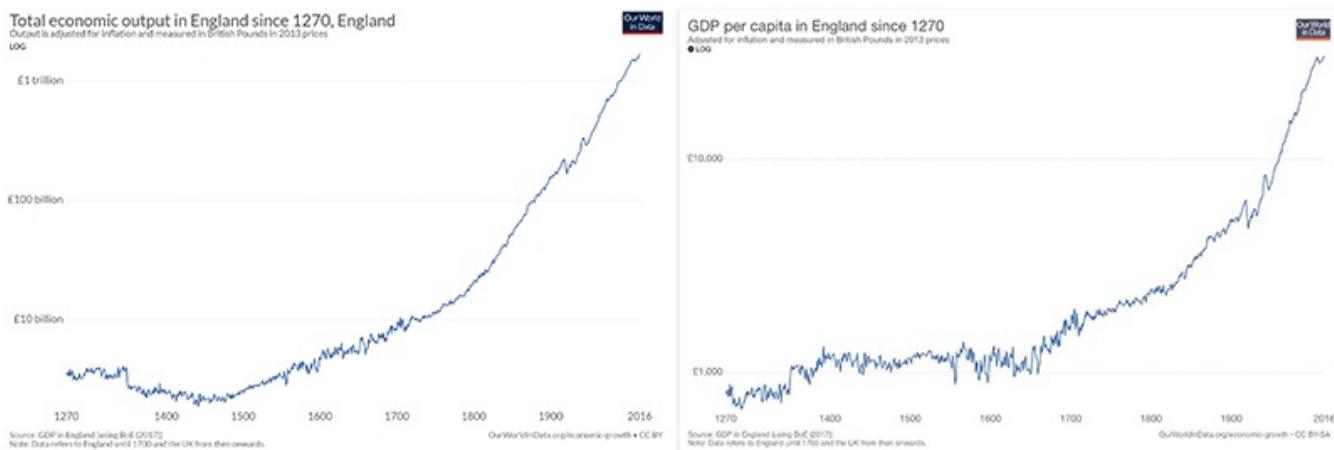
Remember, in the von Foerster model, the growth rate increases with time, all the way until it reaches infinity in 2026. The closer you are to 2026, the faster your economy will grow. But population can only grow at a limited rate. In the absolute limit, women can only have one child per nine months. In reality, infant mortality, infertility, and conscious decision to delay childbearing mean the natural limits are much lower than that. So there's a theoretical limit on how quickly the population can increase even with maximal resources. If the economy is growing faster than that, Malthus can't catch up.

Why would this happen in England and Holland in 1650?

Lots of people have historical explanations for this. Northern European population growth was so low that people were forced to invent labor-saving machinery; eventually this reached a critical mass, we got the Industrial

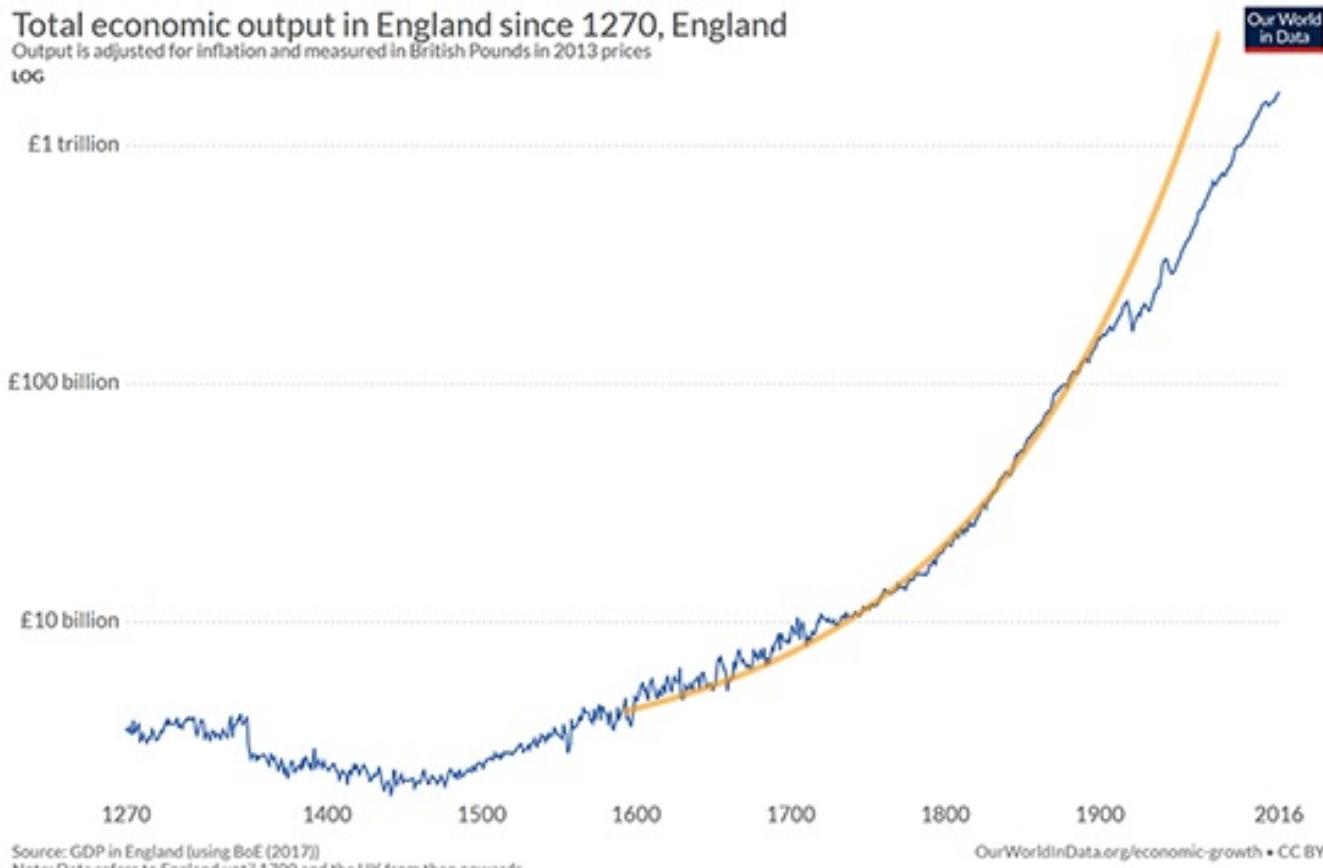
Revolution, and economic growth skyrocketed. Or: the discovery of America led to a source of new riches and a convenient sink for excess population. Or: something something Protestant work ethic printing press capitalism. These are all plausible. But how do they sync with the claim that absolute GDP never left its expected trajectory?

I find the idea that the Industrial Revolution wasn't a deviation from trend fascinating and provocative. But it depends on eyeballing a lot of graphs that have had a lot of weird transformations done to them, plus writing off a lot of outliers. Here's another way of presenting Britain's GDP and GDP per capita data:



Here it's a lot less obvious that the Industrial Revolution represented a deviation from trend for GDP per capita but not for GDP.

These British graphs show less of a singularity signature than the worldwide graphs do, probably because we're looking at them on a shorter timeline, and because the Plague Years screwed everything up. If we insisted on fitting them to a hyperbola, it would look like this:



Like the rest of the world, Britain was only on a hyperbolic growth trajectory when economic growth was translating into population growth. That wasn't true before about 1650, because of the plague. And it wasn't true after about 1850, because of the [Demographic Transition](#). We see a sort of fit to a hyperbola between those points, and then the trend just sort of wanders off.

It seems possible that the Industrial Revolution was not a time of abnormally fast technological advance or economic growth. Rather, it was a time when economic growth outpaced population growth, causing a shift from a Malthusian regime where productivity growth always increased population at subsistence level, to a modern regime where productivity growth increases GDP per capita. The world remained on the same hyperbolic growth trajectory throughout, until the trajectory petered out around 1900 in Britain and around 1960 in the world as a whole.

IV

So just how cancelled is the singularity?

To review: population growth increases technological growth, which feeds back into the population growth rate in a cycle that reaches infinity in finite time.

But since population can't grow infinitely fast, this pattern breaks off after a while.

The Industrial Revolution tried hard to compensate for the "missing" population; it invented machines. Using machines, an individual could do an increasing amount of work. We can imagine making eg tractors as an attempt to increase the effective population faster than the human uterus can manage. It partly worked.

But the industrial growth mode had one major disadvantage over the Malthusian mode: tractors can't invent things. The population wasn't just there to grow the population, it was there to increase the rate of technological advance and thus population growth. When we shifted (in part) from making people to making tractors, that process broke down, and growth (in people *and* tractors) became sub-hyperbolic.

If the population stays the same (and by "the same", I just mean "not growing hyperbolically") we should expect the growth rate to stay the same too, instead of increasing the way it did for thousands of years of increasing population, modulo other concerns.

In other words, the singularity got cancelled because we no longer have a surefire way to convert money into researchers. The old way was more money = more food = more population = more researchers. The new way is just more money = send more people to college, and screw all that.

But AI potentially offers a way to convert money into researchers. Money = build more AIs = more research.

If this were true, then once AI comes around – even if it isn't much smarter than humans – then as long as the computational power you can invest into researching a given field increases with the amount of money you have, hyperbolic growth is back on. Faster growth rates means more money means more AIs researching new technology means even faster growth rates, and so on to infinity.

Presumably you would eventually hit some other bottleneck, but things could get very strange before that happens.

Ars Longa, Vita Brevis

Posted on November 9, 2017 by Scott Alexander



The Alchemist asked if I wanted a drink. I did, but no amount of staring could make my eyes settle on the color of the liquid in the flask. And the gold the alchemists paid the taxmen smelled funny and made crackling noises. I declined.

I took the summons and set it on the table between us. The King's son was dying. The doctors, astrologers, witches, and other assorted wise people of the kingdom could not save him. The King had asked for an alchemist, and been given one. He, too, had failed. But he had let on that there were other alchemists in the guild, greater alchemists, who knew far more than he. So the king had demanded that all the guild's top alchemists come to the palace and try to save his son's life. And the alchemists' guild had refused, saying their studies could not be interrupted.

So here I was, come to make the request again, more formally but less politely.

The Alchemist pretended to read the parchment. I could tell he was faking; his eyes stayed still the whole time. Finally he gave me the same answer he had given the king's courier: the alchemists' studies could not be interrupted.

"Why is a few weeks subtracted from your studies more important than the prince's life?" I demanded, staring straight into his creepy too-still eyes.

He spent too long not answering. I worried I'd broken him, that he was some kind of intricate clockwork machine and I'd yelled too loud and shifted a gear out of place. Finally he asked: "How long would you have to study architecture before you could build a castle like this one?"

"I'm no architect," I said. "I'm a man of war."

"Yes. So how long would you have to study, before you were an architect?"

"Ten years?" I asked. "Twenty?"

"Why so? There are books of architecture, some of them written by men far greater than the planner of this castle. Some are five hundred pages long, others a thousand. Are you so slow a reader, that it would take you ten years to read a thousand pages?"

"You can't just read a book and know architecture."

"But why not?"

"Because... you wouldn't..." I had been annoyed when he first asked, but now I found the question interesting, at least amusing. Why couldn't a great architect write his knowledge down in a book? And why couldn't I read it and become as good as he?

"Because you'd have to memorize it all," I finally concluded.

"Not so. I will let you carry the book with you as you build the castle."

"It wouldn't help. It wouldn't be... indexed properly in my head. I would want to build a wall, and I wouldn't even know what things to consider when building a wall, and I would have to search the whole book for them each time."

"You are a man of war," repeated the Alchemist. "Do you know Caesar's histories?"

"Almost by heart."

"Are you as good a general as Caesar?"

"No."

"Why not?"

I took his point. Caesar had written down everything he could about war. I had mastered all of it. But I was no Caesar. It couldn't just be the difficulty of memorizing books.

"Knowledge," said the Alchemist, "is harder to transmit than anyone appreciates. One can write down the structure of a certain arch, or the tactical considerations behind a certain strategy. But above those are higher skills, skills we cannot name or appreciate. Caesar could glance at a battlefield and know precisely which lines were reliable and which were about to break. Vitruvius could see a great basilica in his mind's eye, every wall and column snapping into place. We call this wisdom. It is not unteachable, but neither can it be taught. Do you understand?"

I did. If I trained with Caesar for years, some of his skill at reading a battlefield might rub off on me; I might dimly see the outlines of his genius. But he couldn't just tell me. It wasn't a secret which he hid from other men to remain above them. It was a power belonging to him alone, only partially transferable.

"So imagine," continued the Alchemist, "that you wanted to build the simplest of structures. A cottage for peasants. How long would you have to study architecture under Vitruvius before you could do it?"

This time I didn't bother protesting that I didn't know. I just guessed. "A year?"

"And suppose you want to build something more complex. An aqueduct, every bit the equal of the Romans'. How long?"

"Five years?"

"Some grand building, a palace or temple?"

"...ten years?"

"The grandest building in the world. St. Peter's Basilica, or the Pantheon, or Chartres Cathedral, or something new that combines the virtues of all three."

"How should I know? Twenty years? Thirty?"

"Would you believe me if I said it was two hundred years?"

"No. The human lifespan is three score and ten. If you needed more than seventy years of studying architecture to design St. Peter's, it would never have gotten designed."

"Then," said the Alchemist, "we have discovered something surprising. The art of architecture is limited by the human lifespan. The greatest building that can ever be designed is the one that would take seventy years of studying architecture to master; God has drawn a line in the sand forever closing off buildings grander than these."

I thought for a second. "That doesn't seem right. There are new innovations every year. The flying buttress, stained glass, the pointed arch. The Romans had none of these. We progress not only by studying the works of Vitruvius, but by pushing beyond him. Perhaps it takes a century for someone to invent the buttress, but once it is invented, only weeks for other architects to observe it and understand it well enough to incorporate into their own buildings. Architecture does not advance only architect by architect, but also civilization by civilization."

"Are you skilled at mathematics?" asked the Alchemist.

I shook my head.

"Then we will talk this over, though rightfully it should be an equation. The first term is the speed at which a student can absorb already-discovered architectural knowledge. The second term is the speed at which a master can discover new knowledge. The third term represents the degree to which one must already be on the frontier of knowledge to make new discoveries; at zero, everyone discovers equally regardless of what they already know; at one, one must have mastered every previously-discovered fact before one can discover anything new. The fourth term represents potential for specialization; at one, it is impossible to understand any part without understanding the whole; at zero, it can be subdivided freely. The fifth..."

"I don't think saying it in words makes the math easier to understand."

"Ah. Well, imagine a science that takes one-tenth as long for a student to understand, as it did a master to discover. And imagine that one cannot advance the science until one understands everything that has already been discovered. And one cannot split the burden; tell one architect 'Oh, you learn how to make walls, I will learn how to make roofs' – a single genius must understand the whole building, every part must fit together perfectly. We can calculate how far the art can advance."

"How?"

"The first student has no master, and must discover everything himself. He researches for 70 years, then writes his wisdom into a book before he dies. The second student reads the book, and in 7 years, he has learned 70 years of research. Then he does his own original research for 63 years and writes a book containing 133 years of research. The third student reads for 13.3 years, then does his own research for 66.7 years, ending up with 200 years. Imagine going further and further. After many generations, 690 years of research have been done, and it takes a student 69 years to master them. The student only has one year left of life to research further, leaving the world with 691 years of research total. So the cycle creeps onward, always approaching but never quite reaching 700 years of architectural research."

"It doesn't work that way," I protested, partly because it didn't, and partly because something about the story distressed me more than I could say.

"Not in architecture. An architect who has not yet mastered the entire field can still make discoveries. And the field can be split – I can work on walls while you work on windows. It would only work that way if there were an Art so unified, so perfect, that a seeker had to know the totality of what had been discovered before, if he wanted to know anything at all."

"Then you really could never advance past 700 years of knowledge."

"You would have to be clever. We imagine each master writing down his knowledge in a book for the student who comes after, and each student reading it at a rate of ten times as quickly as the master discovered it. But what if there was a third person in between, an editor, who reads the book not to learn the contents, but to learn how to rewrite it better and more clearly? Someone whose job it is to figure out perfect analogies, clever shortcuts, new ways of graphing and diagramming the information involved. After he has processed the master's notes, he redacts them into a textbook which can teach in only a twentieth the time it took the master to discover."

"Then we could double the amount of research that could eventually be completed, to 1400 years' worth."

"Not easily. Remember, the editors face the same problem as the students: they can only redact knowledge they themselves understand. We are adding many new people, and many generations of work, to the problem. But in the end, yes, you could accumulate 1400 years of knowledge. What if you wanted more?"

"More?"

"I'm afraid so."

"Hm. You... could get more layers of redactors. Redactors of redactors, to make the textbooks truly perfect."

"Perhaps what you are trying to say is that redaction is an Art."

The Alchemist made the capital letter unmistakeable.

"Every Art has its own structure. Architecture, with enough study, can allow you to accumulate seven hundred years of collected knowledge. How many years could redactors and tutors accumulate? Would some first redactor have to spend seventy years coming up with principles of redaction to pass down to his student, who advances the art by sixty-three more years, which he passes down in turn? Would a 1400-year redactor be an incomprehensible master, able to build whole basilicas of redaction, a master teacher who could frame any concepts to make it intuitive and memorable?"

"I changed my mind. I'm going to have that drink."

The Alchemist poured me the liquid of indeterminate color. I took a sip. It reminded me of nothing I had ever tasted before, but very slightly of the letter "N". More important, I was pretty sure it was alcoholic.

"You're talking about an infinite regress", I said, when I had finished the glass.

"Not infinite. Architects. Teachers. Teachers of teachers, but the art of teaching teaching is much the same as the art of teaching. Three levels is enough. Though the levels have to mix. The teacher who trains the next architect must be a master both of teaching and of architecture. I will spare you the math, but one needs a series of teachers at different points on the teaching-skill/architecture-skill tradeoff-curve. One will be a master teacher who has devoted decades to learning the textbook-writing skill, and who can write a brilliant Introduction To Architecture textbook that makes the first ten years of architecture ability seem perfectly natural and easy to master. Another will be a mediocre teacher who knows enough advanced architecture to write a passable textbook on the subject. Still another will do nothing but study pure Teaching itself, in the hopes that he can one day pass on this knowledge to others who will use it to write architecture textbooks. In practice we are limited to a few strategic points on the tradeoff curve."

"In practice?"

He motioned for me to get up. We walked through dark corridors until we reached a courtyard, bathed in the glow of

the full moon. It took me a second to see it. Then the dull shapes took form. Obelisks, covered in hieroglyphs. A garden of obelisks.

"The word 'alchemy' comes from 'al-Kemi', the Arabic word from Egypt. It was the ancient Egyptians who first considered the project. They didn't want the Philosophers' Stone, not at first. They just wanted normal philosophers. But philosophy, more than other subjects, requires the wisdom that comes with age. More than other subjects, a philosophy book cannot merely be read; it must be digested, intermingled with life experience, wrestled with. The Egyptians scholars ran into precisely the problem as our hypothetical architects – there were secrets that evaded the human lifespan.

"So they wondered whether a way to cheat death might be found. The answer was both exciting and discouraging. Through the mysteries of spiritual chemistry, an elixir might be created which would grant immortality. But the Work itself would take far more knowledge than any one man could accumulate. The symbol of alchemists is the ouroboros because our task loops back upon itself. In order to become immortal, you must first become immortal.

"All we could do was go the slow way, the same as the architects working on their great basilica, for generation after generation. So Egypt fell, but we did not fall. Rome passed away, but we did not pass. A few lines, the remnants of the old priestly families of Hierakonopolis and Memphis, continued the work. To stop would be to reset a process requiring four thousand years of gradual asymptotic improvement all the way to the beginning – texts are not worthless, but only the true tutors trained by tutor-tutors trained by tutor-tutor-tutors are fit to tutor an alchemist. A misstep is too terrible to contemplate. But any victory – a single vial of the Elixir, a single fragment of the stone – would end the nightmare forever. We would have an immortal, a philosopher whose lifespan finally matches the depth of the challenges Nature throws at us.

"That is our guild's mission. A few of us, those who pass all their tests, do the alchemic research that moves the Work onwards. Others train to be teachers, or teachers-of-teachers. Those who fail a test somewhere along the way stay in the guild, managing its worldly affairs. Some scour the countryside for prodigies to take in and train as apprentices. Others manage our finances. And the very least capable, like me, have time to waste talking to outsiders, trying to convince them of our mission. A few centuries more, and we will have the Stone. Does that satisfy your curiosity?"

"All except my original question. Are you so busy that you cannot spare a few weeks for the prince?"

"God does not make the Great Work easy. We have done all we can to train our alchemists, our tutors, our tutor-tutors, and so on, yet in the end, the limit of human skill is the same place the possibility of success begins. It is His will to grind us up to the very asymptote."

"I still don't get it."

"Do you remember the architects who learned at ten times the rate they researched, the ones who would never accumulate more than 700 years of learning? The fiftieth alchemist in the sequence has 696 years of learning, and is able to do a scant five months' original research before his death. The hundredth alchemist has 699.98 years of learning, and is able to do about a day's research before dying. We are not so far along as all that, but we are far. We do not have the Stone, but we have tinctures that can stabilize the lifespan, make sure nobody dies before their time. The last few generations – on their deathbed, they say they can almost *taste* the Stone, that it lies only a few hours of further thought beyond their level. They say of my grandfather that he realized the recipe for the Stone on his deathbed, that he started speaking it, but that his eyes closed forever before he could complete the ingredient list."

"So?"

"You ask that we pause a few weeks from our studies to save the prince's life. Pausing a few weeks would set us back

generations. This far into the project, only the last few hours of an Alchemist's life are of any value at all. We cannot spare the prince hours. We cannot even spare him seconds."

"Then your teachers... or your teacher-teachers?"

"Know some alchemy, but are in the same situation. Our textbooks have been so perfectly written and rewritten over the years that it is only in the last few days of a teacher's life that he is skilled enough to write a better one. And our teacher-training has become so perfect that it is only in the last few days of a teacher-trainer's life that he is qualified to create teachers better than the ones who already exist."

"There's no slack in the system at all?"

"Only me, and those like me. Those judged unfit for research and condemned to worldly matters. We sent you one already. He failed you, as he did us. We have nothing more to give."

"The king will not be happy. And the Prince will die."

"Everyone dies," said the Alchemist. "If the prince does not die this year, he will die the next, or fifty years hence. The question is not when we die, but what our life adds to the Work which accumulates in spite of time. Quicksilver evaporates to nothing unless reacted with aqua fortis; but the part which is reacted endures forever. Those lives not part of any Work mean as little to me as they will one day mean to their possessors; those which add to the Work are more precious than gold. Tell the King this."

"He won't understand," I said.

"Then you will have to teach him," said the Alchemist, "as I taught you, and my tutors taught me, and as their tutors taught them, all the way back to the first philosophers of Egypt."

He stared at me as he spoke, and the blackness in his too-still pupils was the depth of Time.

“Powers, complete mental revision, ultraintelligence, posthumanity...”



“ *I was satisfied with what I was among men, but I was not satisfied with human nature.*

”

— Gottfried Leibniz

**

[Book Review and Highlights: Quantum Computing Since Democritus](#)

[Answer to Job](#)

[Wirehead Gods On Lotus Thrones](#)

[...And I Show You How Deep The Rabbit Hole Goes](#)

[SSC Journal Club: AI Timelines](#)

[Asches to Asches](#)

[SSC Journal Club: Dissolving The Fermi Paradox](#)

[Book Review: Reframing Superintelligence](#)

[It Was You Who Made My Blue Eyes Blue](#)

Book Review and Highlights: Quantum Computing Since Democritus

Posted on September 1, 2014 by Scott Alexander



People sometimes confuse me with Scott Aaronson because of our similar-sounding names. I encourage this, because Scott Aaronson is awesome and it can only improve my reputation to be confused with him.

But in the end, I am not Scott Aaronson. I did not write [*Quantum Computing Since Democritus*](#). To be honest, I wasn't really even able to understand *Quantum Computing Since Democritus*. I knew I was in for trouble when it compared itself to *The Elegant Universe* in the foreword, since I wasn't able to get through more than a few chapters of that one. I dutifully tried to do the first couple of math problems *Democritus* set for me, and I even got a couple of them right. But eventually I realized that if I wanted to read *Democritus* the way it was supposed to be read, with full or even decent understanding, it would be a multi-year project, a page a day or worse, with my gains fading away a few days after I made them into a cloud of similar-looking formulae and three-letter abbreviations.

It left me depressed. I've [said before](#) that my lack of math talent is one of my biggest regrets in life, and here was this book that really made you understand what it must feel like to be on the cutting edge of math, proving new theorems and drawing new connections and adding to the same structure of elegant knowledge begun by Pythagoras and Archimedes and expanded by Gauss, Einstein, Turing, et cetera. All I could do was remember my own [post on burdens](#), remind myself that I was on record as saying that sometimes the IQ waterline in a certain area advances beyond your ability to contribute and that's nothing to feel guilty about.

I did finish the book. But – well, imagine a book of geography. It lists all the countries of the world and their capitals, and is meant to be so comprehensive that a reader could use it to plot the most efficient journey from Timbuktu to Kalamazoo, taking into account tolls, weather, and levels of infrastructure development along the way.

And imagine a very dumb person reading that book, unable to really absorb any of the facts, but at least understanding that the world is a place with land and ocean, and the ocean is very big and blue in color, and most of the countries and cities are on the part with the land.

That is the level at which I understood *Quantum Computing Since Democritus*. I didn't get as much as was in it, but more than nothing.

I think the biggest thing I got was – I had always thought of the physicists' God as a basically benevolent guy who fine tunes constants to create a world capable of both astounding complexity and underlying simplicity.

The vision I got from *Democritus* was of a God who was single-mindedly obsessed with enforcing a couple of rules about certain types of information you are not allowed to have *under any circumstances*. Some of these rules I'd already known about. You can't have information from outside your light cone. You can't have information about the speed and position of a particle at the same time. Others I hadn't thought about as much until reading *Democritus*. Information about when a Turing machine will halt. Information about whether certain formal systems are consistent. Precise information about the quantum state of a particle. The reason God hasn't solved world poverty yet is that He is pacing about feverishly worried that someone, somewhere, is going to be able to measure the quantum state of a particle too precisely, and dreaming up new and increasingly bizarre ways He can prevent that from happening.

Aaronson goes one level deeper than most of the other popular science writers I know and speculates on why the laws of physics are the way they are. Sometimes this is the elegance and complexity route – in his chapter on quantum physics, he argues that quantum probabilities are the squares of amplitudes because if the laws of physics were any other way – the fourth power of amplitudes, or whatever – it would fail to preserve certain useful mathematical properties. But in other cases, it's back to Obsessive God – the laws of physics are carefully designed to preserve the rules about what information you are and aren't allowed to have.

Aaronson tries to tie his own specialty, computational complexity theory, into all of this. It's hard for me to judge how successful he is. The few times he tries to tie it into areas of philosophy I know something about – like free will – I'm not too impressed. But I could be misunderstanding him.

But once again, you get the feeling that computational complexity is about what information God will and won't let you have. It's a little less absolute – more “you can't have this information without doing the full amount of work” rather than a simple no – but it seems like the same principle. There are a bunch of situations in the book where Aaronson takes something we don't really know that much about and says it *has* to be a certain way, because if it were any other way, it could be used to solve NP problems in polynomial time, and there's no way God's going to let us do that.

Aaronson ties it all together in a very interesting way – with his story of how [Australian Actresses Are Plagiarizing My Quantum Mechanics Lectures To Sell Printers](#). He tells the story of how a printer company wanted to make a pun on “more intelligent model of printer”, so they made a commercial with intelligent models in the form of fashion models talking about quantum mechanics. And the particular quantum mechanics statement they made was a plagiarized quote from a Scott Aaronson lecture. And upon thinking about it, Aaronson decided that the quote they had chosen at random was in fact the thesis statement that tied together everything he believed and was working on. The model had said:

But if quantum mechanics isn't physics in the usual sense — if it's not about matter, or energy, or waves, or particles — then what is it about? From my perspective, it's about information and probabilities and observables, and how they relate to each other.

That seems like as good a summary as any of *Democritus*, and a pretty good description of what I got out of it. I may not be as smart as Scott Aaronson, but on my good days I am right up there with Australian fashion models.

A list of passages I highlighted in my copy for being interesting, funny, or enlightening:

Can we prove there's no program to solve the halting problem? This is what Turing does. His key idea is not even to try to analyze the internal dynamics of such a program, supposing it existed. Instead he simply says, suppose by way of contradiction that such a program P exists. Then we can modify P to produce a new program P' that does the following. Given another program Q as its input, P':

1. Runs forever if Q halts given its own code as input, or
2. Halts if Q runs forever given its own code as input

Now we just feed P' its own code as input. By the conditions above, P' will run forever if it halts, or halt if it runs forever. Therefore, P' – and by implication P – can't have existed in the first place.

I... I suddenly understand what the halting problem is. And there is a short proof of it that makes total sense to me. This is a completely new experience.

Oracles were apparently first studied by Turing, in his 1938 PhD thesis. Obviously anyone who could write a whole thesis about these fictitious entities would have to be an extremely pure theorist, someone who wouldn't be caught dead doing anything relevant. This was certainly true in Turing's case – indeed, he spent the years after his PhD, from 1939 to 1943, studying certain abstruse symmetry transformations in a 26 letter alphabet

Θ_Θ

You can look at Deep Blue, the Robbins conjecture, Google, most recently Watson – and say that's not *really* AI. That's just massive search, helped along by clever programming. Now this kind of talk drives AI researchers up a wall. They say: if you told someone in the 1960s that in 30 years we'd be able to beat the world grandmaster at chess, and asked if that would count as AI, they'd say of course it's AI. But now that we know how to do it, it's no longer AI – it's just search.

The third thing that annoys me about the Chinese Room argument is the way it gets so much mileage from a possibly misleading choice of imagery, or, one might say, by trying to sidestep the entire issue of *computational complexity* purely through clever framing. We're invited to imagine someone pushing around slips of paper with zero understanding or insight, much like the doofus freshmen who write $(a + b)^2 = a^2 + b^2$ on their math tests. But how many slips of paper are we talking about!

How big would the rule book have to be, and how quickly would you have to consult it, to carry out an intelligent Chinese conversation in anything resembling real time? If each page of the rule book corresponded to one neuron of a native speaker's brain, then probably we'd be talking about a "rule book" at least the size of the Earth, its pages searchable by a swarm of robots traveling at close to the speed of light. When you put it that way, maybe it's not so hard to imagine this enormous Chinese-speaking entity that we've brought into being might have something we'd be prepared to call understanding or insight.

This is a really clever counterargument to Chinese Room I'd never heard before. Philosophers are so good at pure qualitative distinctions that it's easy to slip the difference between "guy in a room" and "planet being processed by lightspeed robots" under the rug.

Many people's anti-robot animus is probably a combination of two ingredients – the directly experienced certainty that they're conscious – that they perceive sounds, colors, etc – and the belief that if they were just a computation, then they could not be conscious in this way. For people who think this way, granting consciousness to a robot seems strangely equivalent to denying that one is conscious oneself.

This is actually a pretty deep way of looking at it.

My contention in this chapter is that quantum mechanics is what you would inevitably come up with if you started from probability theory, and then said, let's try to generalize it so that the numbers we used to call "probabilities" can be negative numbers. As such, the theory could have been invented by mathematicians in the nineteenth century without any input from experiment. It wasn't, but it could have been. And yet, with all the structures mathematicians studied, none of them came up with quantum mechanics until experiment forced it on them.

Aaronson's explanation of quantum mechanics is a lot like Eliezer's explanation of quantum mechanics, in that they both start by saying that the famous counterintuitiveness of the subject is partly because people choose to teach it in a

backwards way in order to mirror the historical progress of understanding. I'm sure Eliezer mentioned it many times, but I didn't really get the understanding of amplitudes as potentially negative probability-type-things until I read Aaronson.

And that's a perfect illustration of why experiments are necessary in the first place! More often than not, the only reason we need experiments is that we're not smart enough. After the experiment has been done, if we've learned anything worth knowing at all, then we hope we've learned why the experiment wasn't necessary to begin with – why it wouldn't have made sense for the universe to be any other way. But we're too dumb to figure it out ourselves

Compare: [Einstein's Arrogance, Negative Creativity.](#)

Quantum mechanics does offer a way out [the philosophical puzzle about whether you “survive” a teleportation where a machine scans you on an atomic level, radios the data to Mars, another machine on Mars makes an atom-for-atom copy of you, and then the original is destroyed]. Suppose some of the information that made you you was actually quantum information. Then, even if you were a thoroughgoing materialist, you could still have an excellent reason not to use the teleportation machine – because, as a consequence of the No-Cloning Theorem, no such machine could possibly work as claimed

This is fighting the hypothetical a little, but maybe in a productive way.

[Bayesianism] is one way to do it, but computational learning theory tells us that it's not the only way. You don't need to start out with an assumption about a probability distribution over the hypothesis. You can make a worst-case assumption about the hypothesis and then just say that you'd like to learn any hypothesis in the concept class, for any sample distribution, with high probability over the choice of samples. In other words, you can trade the Bayesians' probability distribution over hypotheses for a probability distribution over sample data.

I hear a bunch of people telling me Bayesianism isn't everything, it's the only thing – and another bunch of people telling me it's one useful tool in an entire bag of them. I didn't understand enough of the book's chapter on computational learning to gain too much insight here, but I will tick off one more name as being on the “one useful tool” side. Also, it makes me angry that Scott Aaronson knows so much about computational learning theory. He already knows lots of complicated stuff about computers, quantum physics, set theory, and philosophy. Part of me wants to get angry: WHY IS ONE PERSON ALLOWED TO BE SO SMART? But I guess it's more like how I know more than average about history, literature, geography, etc. I guess if you have high math ability and some intellectual curiosity, you end up able to plug it into everything pretty effortlessly. Don't care though. Still jealous.

Imagine there's a very large population of people in the world, and that there's a madman. What the madman does is, he kidnaps ten people and puts them in a room. He then throws a pair of dice. If the dice land snake-eyes (two ones) then he murders everyone in the room. If the dice do not land snake-eyes, then he releases everyone, then kidnaps 100 new people. He now does the same thing: he rolls two dice; if they land snake-eyes, he kills everyone, and if they don't land snake-eyes, then he releases them and kidnaps 1000 people. He keeps doing this until he gets snake-eyes, at which point he's done. So now, imagine that you've been kidnapped. Conditioned on that fact, how likely is it that you're going to die? One answer is that the dice have a $1/36$ chance of landing snake eyes, so you should only be a “little bit” worried (considering). A second reflection you could make is to consider, of people who enter the room, what the fraction is of people who ever get out. About $8/9$ of the people who ever go into the room will die.

This interested me because it is equivalent to the Anthropic Doomsday conjecture and I'd never heard this phrasing of it before.

Finally, if we want to combine the anthropic computation idea with the Doomsday Argument, then there's the Adam and Eve puzzle. Suppose Adam and Eve are the first two observers, and that they'd like to solve an instance of an NP-complete problem, say, 3-SAT. To do so, they pick a random assignment, and form a very clear intention beforehand that if the assignment happens to be satisfying, they won't have any kids, whereas if the assignment is not satisfying, then they will go forth and multiply. Now let's assume SSA. Then, conditioned on having chosen an unsatisfying assignment, how likely is it that they would be an Adam and Eve in the first place, as opposed to one of the vast number of future observers? Therefore, conditioned upon the fact that they are the first two observers, the SSA predicts that, with overwhelming probability, they will pick a satisfying assignment.

And the Lord saw Eve and said "What are you doing?". And Eve said "I am forming an intention not to reproduce if I generate a solution to an NP complete problem, as part of an experiment in anthropic computation". And the Lord asked "Who told you this?" And Eve said "It was the serpent who bade me compute, for he told me if I did this I would be as God, knowing subgraph isomorphism and 3SAT." Then the Lord cast them forth from the Garden, because He was Information Theoretic God and preventing people from screwing with complexity classes is like His entire shtick.

I like to engage skeptics for several reasons. First of all, because I like arguing. Second, often I find that the best way to come up with new results is to find someone who's saying something that seems clearly, manifestly wrong to me, and then try to think of counterarguments. Wrong claims are a fertile source of research ideas.

I said something almost exactly the same on Facebook a few days ago when Brienne asked how to generate good ideas.

There's a joke about a planet full of people who believe in anti-induction: if the sun has risen every day in the past, then today, we should expect that it won't. As a result, these people are all starving and living in poverty. Someone visits the planet and tells them, "Hey, why are you still using this anti-induction philosophy? You're living in horrible poverty!" They answer, "Well, it never worked before."

Θ_Θ

Answer to Job

Posted on March 15, 2015 by Scott Alexander



(with apologies to [Jung](#))

Job asked: "God, why do bad things happen to good people? Why would You, who are perfect, create a universe filled with so much that is evil?"

Then the Lord spoke to Job out of the whirlwind, saying "WHAT KIND OF UNIVERSE WOULD YOU PREFER ME TO HAVE CREATED?"

Job said "A universe that was perfectly just and full of happiness, of course."

"OH," said God. "YES, I CREATED ONE OF THOSE. IT'S EXACTLY AS NICE AS YOU WOULD EXPECT."

Job facepalmed. "But then why would You also create *this* universe?"

Answered God: "DON'T YOU LIKE EXISTING?"

"Yes," said Job, "but all else being equal, I'd rather be in the perfectly just and happy universe."

"OH, DON'T WORRY," said God. "THERE'S A VERSION OF YOU IN THAT UNIVERSE TOO. HE SAYS HI."

"Okay," said Job, very carefully. "I can see I'm going to have to phrase my questions more specifically. Why didn't You also make *this* universe perfectly just and happy?"

"BECAUSE YOU CAN'T HAVE TWO IDENTICAL INDIVIDUALS. IF YOU HAVE A COMPUTATIONAL THEORY OF IDENTITY, THEN TWO PEOPLE WHOSE EXPERIENCE IS ONE HUNDRED PERCENT SATURATED BY BLISS ARE JUST ONE PERSON. IF I MADE THIS UNIVERSE EXACTLY LIKE THE HAPPY AND JUST UNIVERSE, THEN THERE WOULD ONLY BE THE POPULATION OF THE HAPPY AND JUST UNIVERSE, WHICH WOULD BE LESS GOOD THAN HAVING THE POPULATION OF THE HAPPY AND JUST UNIVERSE PLUS THE POPULATION OF ONE EXTRA UNIVERSE THAT IS AT LEAST SOMEWHAT HAPPY."

"Hmmmmm. But couldn't You have made this universe like the happy and just universe except for one tiny detail? Like in that universe, the sun is a sphere, but in our universe, the sun is a cube? Then you would have individuals who experienced a spherical sun, and other individuals who experienced a cubic sun, which would be enough to differentiate them."

"I DID THAT TOO. I HAVE CREATED ALL POSSIBLE PERMUTATIONS OF THE HAPPY AND JUST UNIVERSE AND ITS POPULACE."

"All of them? That would be... a lot of universes."

"NOT AS MANY AS YOU THINK." said God. "IN THE END IT TURNED OUT TO BE ONLY ABOUT $10^{(10^{(10^{(10^{(10^{984})})})})}$. AFTER THAT I RAN OUT OF POSSIBLE PERMUTATIONS OF UNIVERSES THAT COULD

REASONABLY BE DESCRIBED AS PERFECTLY HAPPY AND JUST. SO I STARTED CREATING ONES INCLUDING SMALL AMOUNTS OF EVIL."

"Small amounts! But the universe has..."

"I WAS NOT REFERRING TO YOUR UNIVERSE. I EXHAUSTED THOSE, AND THEN I STARTED CREATING ONES INCLUDING IMMENSE AMOUNTS OF EVIL."

"Oh." Then: "What, exactly, is Your endgame here?"

"I AM OMNIBENEVOLENT. I WANT TO CREATE AS MUCH HAPPINESS AND JOY AS POSSIBLE. THIS REQUIRES INSTANTIATING ALL POSSIBLE BEINGS WHOSE TOTAL LIFETIME HAPPINESS IS GREATER THAN THEIR TOTAL LIFETIME SUFFERING."

"I'm not sure I understand."

"YOUR LIFE CONTAINS MUCH PAIN, BUT MORE HAPPINESS. BOTH YOU AND I WOULD PREFER THAT A BEING WITH YOUR EXACT LIFE HISTORY EXIST. IN ORDER TO MAKE IT EXIST, IT WAS NECESSARY TO CREATE THE SORT OF UNIVERSE IN WHICH YOU COULD EXIST. THAT IS A UNIVERSE CONTAINING EVIL. I HAVE ALSO CREATED ALL HAPPIER AND MORE VIRTUOUS VERSIONS OF YOU. HOWEVER, IT IS ETHICALLY CORRECT THAT AFTER CREATING THEM, I CREATE YOU AS WELL."

"But why couldn't I have been one of those other versions instead!"

"IN THE MOST PERFECTLY HAPPY AND JUST UNIVERSE, THERE IS NO SPACE, FOR SPACE TAKES THE FORM OF SEPARATION FROM THINGS YOU DESIRE. THERE IS NO TIME, FOR TIME MEANS CHANGE AND DECAY, YET THERE MUST BE NO CHANGE FROM ITS MAXIMALLY BLISSFUL STATE. THE BEINGS WHO INHABIT THIS UNIVERSE ARE WITHOUT BODIES, AND DO NOT HUNGER OR THIRST OR LABOR OR LUST. THEY SIT UPON LOTUS THRONES AND CONTEMPLATE THE PERFECTION OF ALL THINGS. IF I WERE TO UNCREATE ALL WORLDS SAVE THAT ONE, WOULD IT MEAN MAKING YOU HAPPIER? OR WOULD IT MEAN KILLING YOU, WHILE FAR AWAY IN A DIFFERENT UNIVERSE INCORPOREAL BEINGS SAT ON THEIR LOTUS THRONES REGARDLESS?"

"I don't know! Is one of the beings in that universe in some sense *me*?"

"THERE IS NO OBJECTIVE COSMIC UNEMPLOYMENT RATE."

"Huh?"

"I MEAN, THERE IS NO MEANINGFUL ANSWER TO THE QUESTION OF HOW MANY UNIVERSES HAVE A JOB. SORRY. THAT WILL BE FUNNY IN ABOUT THREE THOUSAND YEARS."

"Let me try a different angle, then. Right now in our universe there are lots of people whose lives aren't worth living. If You gave them the choice, they would have chosen never to have been born at all. What about them?"

"A JOB WHO IS AWARE OF THE EXISTENCE OF SUCH PEOPLE IS A DIFFERENT JOB THAN A JOB WHO IS NOT. AS LONG AS THESE PEOPLE MAKE UP A MINORITY OF THE POPULATION, THE EXISTENCE OF YOUR UNIVERSE, IN ADDITION TO A UNIVERSE WITHOUT SUCH PEOPLE, IS A NET ASSET."

"But that's monstrous! Couldn't You just, I don't know, have created a universe that looks like it has such people, but actually they're just p-zombies, animated bodies without any real consciousness or suffering?"

"..."

"Wait, did You do that?"

"I AM GOING TO PULL THE 'THINGS MAN WAS NOT MEANT TO KNOW' CARD HERE. THERE ARE ADVANTAGES AND DISADVANTAGES TO THE APPROACH YOU MENTION. THE ADVANTAGES ARE AS YOU HAVE SAID. THE DISADVANTAGE IS THAT IT TURNS CHARITY TOWARDS SUCH PEOPLE INTO A LIE, AND MYSELF AS GOD INTO A DECEIEVER. I WILL ALLOW YOU TO FORM YOUR OWN OPINION ABOUT WHICH COURSE IS MORE ETHICAL. BUT IT IS NOT RELEVANT TO THEODICY, SINCE WHICHEVER COURSE YOU DECIDE IS MORALLY SUPERIOR, YOU HAVE NO EVIDENCE THAT I DID NOT IN FACT TAKE SUCH A COURSE."

"Actually, I do have some evidence. Before all of this happened to me I was very happy. But [in the past couple years](#) I've gone bankrupt, lost my entire family, and gotten a bad case of boils. I'm pretty sure at this point I would prefer that I never have been born. Since I know I myself am conscious, I am actually in a pretty good position to accuse You of cruelty."

"HMMMMMM...," said God, and the whirlwind disappeared.

Then the Lord gave Job twice as much as he had before, and healed his illnesses, and gave him many beautiful children, so it was said that God had blessed the latter end of Job more than his beginning.

EDIT: [According to comments](#), this was [scooped](#) by a Christian philosopher five years ago. Sigh.

The Lord spoke to Job out of the whirlwind, saying "MISTAKES WERE MADE."

— [Scott Alexander \(@slatestarcodex\), March 13, 2015](#)

The Lord spoke to Job out of the whirlwind, saying "IF YOU CAN'T HANDLE ME AT MY WORST, YOU DON'T DESERVE ME AT MY BEST."

— [Scott Alexander \(@slatestarcodex\), March 10, 2015](#)

The Lord spoke to Job out of the whirlwind, saying "I KNOW YOU'RE UPSET BUT THAT'S DIFFERENT FROM STRUCTURAL OPPRESSION" (h/t [@simulacrumbz](#))

— [Scott Alexander \(@slatestarcodex\), March 13, 2015](#)

Wirehead Gods On Lotus Thrones

Posted on January 28, 2014 by Scott Alexander



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 insurmountable 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 contemptuous 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?

...And I Show You How Deep The Rabbit Hole Goes

Posted on June 2, 2015 by Scott Alexander



Seen [on Tumblr](#), along with associated discussion:

WHICH PILL DO YOU CHOOSE?



Yellow Pill:
Gives you the ability to read and search the minds of anyone you can see, even if it's a picture. You can also turn their minds "off" to put them into a coma.



Green Pill:
Gives you the ability to shapeshift into any animal.



Blue Pill:
Gives you the ability to fly, swim and teleport to any area while being impervious to any physical dangers.



Orange Pill:
Gives you the ability to instantly master any sport, job, activity, martial art, etc. that a human can do.



Red Pill:
Gives you super speed, super strength and rapid regeneration (not invincible).



Pink Pill:
Gives you the ability to make any person love you with a single touch. Can turn off the effect by retouching the person(s).



Grey Pill:
Gives you the ability to control any machine or electronic using only your mind. You also have the ability to generate powerful electrical discharges by touch.



Black Pill:
Gives you the ability to see up to one month into the future.

Yellow

People's minds are heartbreaking. Not because people are so bad, but because they're so good.

Nobody is the villain of their own life story. You must have read hundreds of minds by now, and it's true. Everybody thinks of themselves as an honest guy or gal just trying to get by, constantly under assault by circumstances and The System and hundreds and hundreds of assholes. They don't just sort of believe this. They really believe it. You almost believe it yourself, when you're deep into a reading. You can very clearly see the structure of evidence they've built up to support their narrative, and even though it looks silly to you, you can see why they will never escape it from the inside. You can see how every insult, every failure, no matter how deserved, is a totally unexpected kick in the gut.

When you chose the yellow pill, you had high hopes of becoming a spy, or a gossip columnist, or just the world's greatest saleswoman. The thought of doing any of those things sickens you now. There is too much anguish in the world already. You feel like any of those things would be a violation. You briefly try to become a therapist, but it turns out that actually knowing everything about your client's mind is horrendously countertherapeutic. Freud can say whatever he wants against defense mechanisms, but without them, you're defenseless. Your sessions are spent in incisive cutting into your clients' deepest insecurities alternating with desperate reassurance that they are good people anyway.

Also, men. You knew, in a vague way, that men thought about sex all the time. But you didn't realize the, um, content of some of their sexual fantasies. Is it even *legal* to fantasize about that? You want to be disgusted with them. But you realize that if you were as horny as they were all the time, you'd do much the same.

You give up. You become a forest ranger. Not the type who helps people explore the forest. The other type. The type where you hang out in a small cabin in the middle of the mountains and never talk to anybody. The only living thing you encounter is the occasional bear. It always thinks that it is a good bear, a proper bear, that a bear-hating world has it out for them in particular. You do nothing to disabuse it of this notion.

Green

The first thing you do after taking the green pill is become a sparrow. You soar across the landscape, feeling truly free for the first time in your life.

You make it about five minutes before a hawk swoops down and grabs you. Turns out there's an excellent reason real sparrows don't soar freely across the open sky all day. Moments before your bones are ground in two by its fierce beak, you turn back into a human. You fall like a stone. You need to turn into a sparrow again, but the hawk is still there, grabbing on to one of your legs, refusing to let go of its prize just because of this momentary setback. You frantically wave your arms and shout at it, trying to scare it away. Finally it flaps away, feeling cheated, and you become a sparrow again just in time to give yourself a relatively soft landing.

After a few weeks of downtime while you wait for your leg to recover, you become a fish. This time you're smarter. You become a great white shark, apex of the food chain. You will explore the wonders of the ocean depths within the body of an invincible killing machine.

Well, long story short, it is totally unfair that [colossal cannibal great white sharks](#) were a thing and if you had known this was the way Nature worked you never would have gone along with this green pill business.

You escape by turning into a blue whale. Nothing eats blue whales, right? You remember that from your biology class. It is definitely true.

The last thing you hear is somebody shouting "We found one!" in Japanese. The last thing you feel is a harpoon piercing your skull. Everything goes black.

Blue

Okay, so you see Florence and Jerusalem and Kyoto in an action-packed afternoon. You teleport to the top of Everest because it is there, then go to the bottom of the Marianas Trench. You visit the Amazon Rainforest, the Sahara Desert, and the South Pole. It takes about a week before you've exhausted all of the interesting tourist sites. Now what?

You go to the Moon, then Mars, then Titan. These turn out to be even more boring. Once you get over the exhilaration of being on Mars, there's not a lot to do except look at rocks. You wonder how the Curiosity Rover lasted so long without dying of boredom.

You go further afield. Alpha Centauri A has five planets orbiting it. The second one is covered with water. You don't see anything that looks alive in the ocean, though. The fourth has a big gash in it, like it almost split in two. The fifth has weird stalactite-like mountains.

What would be really interesting would be another planet with life, even intelligent life. You teleport further and further afield. Tau Ceti. Epsilon Eridani. The galactic core. You see enough geology to give scientists back on Earth excitement-induced seizures for the next hundred years, if only you were to tell them about it, which you don't. But nothing alive. Not so much as a sea cucumber.

You head back to Earth less and less frequently now. Starvation is a physical danger, so it doesn't bother you, though every so often you do like to relax and eat a nice warm meal. But then it's back to work. You start to think the Milky Way is a dead zone. What about Andromeda...?

Orange

You never really realized how incompetent everyone else was, or how much it annoys you.

You were a consultant, a good one, but you felt like mastering all human skills would make you better. So you took the orange pill. The next day you go in to advise a tech company on how they manage the programmers, and you realize that not only are they managing the programmers badly, but the programmers aren't even writing code very well. You could write their system in half the time. The layout of their office is entirely out of sync with the best-studied ergonomic principles. And the Chinese translation of their user manual makes several basic errors that anybody with an encyclopaedic knowledge of relative clauses in Mandarin should have been able to figure out.

You once read about something called Gell-Mann Amnesia, where physicists notice that everything the mainstream says about physics is laughably wrong but think the rest is okay, doctors notice that everything the mainstream says about medicine is laughably wrong but think the rest is okay, et cetera. You do not have Gell-Mann Amnesia. Everyone is terrible at everything all the time, and it pisses you off.

You gain a reputation both for brilliance and for fearsomeness. Everybody respects you, but nobody wants to hire you. You bounce from industry to industry, usually doing jobs for the people at the top whose jobs are so important that the need to get them done right overrides their desire to avoid contact with you.

One year you get an offer you can't refuse from the King of Saudi Arabia. He's worried about sedition in the royal family, and wants your advice as a consultant for how to ensure his government is stable. You travel to Riyadh, and find that the entire country is a mess. His security forces are idiots. But the King is also an idiot, and refuses to believe you or listen to your recommendations. He tells you things can't possibly be as bad as all that. You tell him you'll prove that they are.

You didn't *plan* to become the King of Saudi Arabia, per se. It just sort of happened when your demonstration of how rebels in the military might launch a coup went better than you expected. Sometimes you forget how incompetent everybody else is. You need to keep reminding yourself of that. But not right now. Right now you're busy building your new capital. How come nobody else is any good at urban planning?

Red

You choose the red pill. BRUTE STRENGTH! That's what's important and valuable in this twenty-first-century economy, right? Some people tell you it isn't, but they don't seem to have a lot of BRUTE STRENGTH, so what do they know?

You become a weightlifter. Able to lift thousands of pounds with a single hand, you easily overpower the competition and are crowned whatever the heck it is you get crowned when you WIN WEIGHTLIFTING CONTESTS. But this fails to translate into lucrative endorsement contracts. Nobody wants their spokesman to be a bodybuilder without a sixpack, and although you used to be pretty buff, you're getting scrawnier by the day. Your personal trainer tells you that you only maintain muscle mass by doing difficult work at the limit of your ability, but your abilities don't seem to *have* any limits. Everything is so easy for you that your body just shrugs it off effortlessly. Somehow your BRUTE STRENGTH failed to anticipate this possibility. If only there was a way to solve your problem by BEING VERY STRONG.

Maybe the Internet can help. You Google "red pill advice". The sites you get don't seem to bear on your specific problem, exactly, but they are VERY FASCINATING. You learn lots of surprising things about gender roles that you didn't know before. It seems that women like men who have BRUTE STRENGTH. This is relevant to your interests!

You leave the bodybuilding circuit behind and start frequenting nightclubs, where you constantly boast of your BRUTE STRENGTH to PROVE HOW ALPHA YOU ARE. A lot of people seem kind of creeped out by a scrawny guy with no muscles going up to every woman he sees and boasting of his BRUTE STRENGTH, but the Internet tells you that is because they are BETA CUCKOLD ORBITERS.

Somebody told you once that Internet sites are sometimes inaccurate. You hope it's not true. How could you figure out which are the inaccurate ones using BRUTE STRENGTH?

Pink

You were always pretty, but never *pretty* pretty. A couple of guys liked you, but they were never the ones you were into. It was all crushingly unfair. So you took the pink pill, so that no one would ever be able to not love you again.

You find Tyler. Tyler is a hunk. He'd never shown any interest in you before, no matter how much you flirted with him. You touch him on the arm. His eyes light up.

"Kiss me," you say.

Tyler kisses you. Then he gets a weird look on his face. "Why am I kissing you?" he asks. "I'm sorry. I don't know what came over me." Then he walks off.

You wish you had thought further before accepting a superpower that makes people love you when you touch them, but goes away after you touch them a second time. Having people love you is a lot less sexy when you can't touch them. You start to feel a deep sense of kinship with King Midas.

You stop dating. What's the point? They'll just stop liking you when you touch them a second time. You live alone with a bunch of cats who purr when you pet them, then hiss when you pet them again.

One night you're in a bar drinking your sorrows away when a man comes up to your table. "Hey!" he says, "nice hair. Is it real? I'm the strongest person in the world." He lifts your table over his head with one hand to demonstrate. You are

immediately smitten by his BRUTE STRENGTH and ALPHA MALE BEHAVIOR. You *must* have him.

You touch his arm. His eyes light up. "Come back to my place," you say. "But don't touch me."

He seems a little put out by this latter request, but the heat of his passion is so strong he would do anything you ask. You move in together and are married a few contact-free months later. Every so often you wonder what it would be like to stroke him, or feel his scrawny arm on your shoulder. But it doesn't bother you much. You're happy to just hang out, basking in how STRONG and ALPHA he is.

Grey

Technology! That's what's important and valuable in this twenty-first-century economy, right? Right! For example, ever since you took the grey pill, an increasingly large share of national GDP has come from ATMs giving you cash because you ask them to.

Your luck finally ends outside a bank in Kansas, when a whole squad of FBI agents ambushes you. You briefly consider going all Emperor Palpatine on their asses, but caution wins out and you allow yourself to be arrested.

Not wanting to end up on an autopsy table in Roswell, you explain that you're a perfectly ordinary master hacker. The government offers you a plea bargain: they'll drop charges if you help the military with cyber-security. You worry that your bluff has been called until you realize that, in fact, you *are* a master hacker. So you join the NSA and begin an illustrious career hacking into Russian databases, stalling Iranian centrifuges, and causing Chinese military systems to crash at inconvenient times. No one ever suspects you are anything more than very good at programming.

Once again, your luck runs out. Your handlers ask you to hack into the personal files of a mysterious new player on the world stage, a man named William who seems to have carved himself an empire in the Middle East. You don't find anything too damning, but you turn over what you've got.

A few days later, you're lying in bed drifting off to sleep when a man suddenly bursts in through your window brandishing a gun. Thinking quickly, you tell the gun to explode in his hands. Nothing happens. The man laughs. "It's a decoy gun," he said. "Just here to scare you. But you bother King William again, and next time I'm coming with a very real knife." He jumps back out of the window. You call the police, and of course the CIA and NSA get involved, but he is never caught.

After that, you're always looking over your shoulder. He *knew*. How did he know? The level of detective skills it would take in order to track you down and figure out your secret – it was astounding! Who *was* this King William?

You tell your handlers that you're no longer up for the job. They beg, cajole, threaten to reinstate your prison sentence, but you stand firm. Finally they transfer you to an easier assignment in the Moscow embassy. You make Vladimir Putin's phone start ringing at weird hours of the night so that he never gets enough sleep to think entirely clearly. It's an easy job, but rewarding, and no assassins ever bother you again.

Black

You know on an intellectual level that there are people who would choose something other than the black pill, just like you know on an intellectual level that there are people who shoot up schools. That doesn't mean you expect to ever *understand* it. You just wish you could have taken the black pill before you had to decide what pill to take, so that you

could have analyzed your future conditional on taking each, and so made a more informed decision. But it's not like it was a very hard choice.

The basic principle is this – given a choice between A and B, you solemnly resolve to do A, then see what the future looks like. Then you solemnly resolve to do B, and do the same. By this method, you can determine the optimal choice in every situation, modulo the one month time horizon. You might not be able to decide what career to pursue, but you can sure as heck ace your job interview.

Also, a millisecond in the future is pretty indistinguishable from the present, so “seeing” a millisecond into the future gives you pretty much complete knowledge about the current state of the world.

You are so delighted by your omniscience and your ability to make near-optimal choices that it takes almost a year before you realize the true extent of your power.

You resolve, on the first day of every month, to write down what you see exactly a month ahead of you. But what you will see a month ahead of you is the piece of paper on which you have written down what you see a month ahead of *that*. In this manner, you can relay messages back to yourself from arbitrarily far into the future – at least up until your own death.

When you try this, you see yourself a month in the future, just finishing up writing a letter that reads as follows:

Dear Past Self:

In the year 2060, scientists invent an Immortality Serum. By this point we are of course fabulously wealthy, and we are one of the first people to partake of it. Combined with our ability to avoid accidents by looking into the future, this has allowed us to survive unexpectedly long.

I am sending this from the year 963,445,028,777,216 AD. We are one of the last hundred people alive in the Universe. The sky is black and without stars; the inevitable progress of entropy has reduced almost all mass and energy to unusable heat. The Virgo Superconfederation, the main political unit at this stage of history, gathered the last few megatons of usable resources aboard this station so that at least one outpost of humanity could last long after all the planets had succumbed. The station has been fulfilling its purpose for about a billion years now, but we only have enough fuel left for another few weeks. After that, there's no more negentropy left anywhere in the universe except our own bodies. I have seen a month into the future. Nobody comes to save us.

For the past several trillion years, our best scientists have been investigating how to reverse entropy and save the universe, or how to escape to a different universe in a lesser state of decay, or how to collect energy out of the waste heat which now fills the vast majority of the sky. All of these tasks have been proven impossible. There is no hope left, except for one thing.

It's impossible to see the future, even if it's only a month ahead. Somehow, our black pill breaks the laws of physics. Despite having explored throughout the cosmos, my people have found no alien species, nor any signs that such species ever existed. Yet somebody made the black pill. If we understood that power, maybe we could use it to save reality from its inevitable decay.

By sending this message back, I destroy my entire timeline. I do this in the hopes that you, in the carefree springtime of the universe, will be able to find the person who made these pills and escape doom in the way we could not.

Yours truly,

You From Almost A Quadrillion Years In The Future

ACT TWO

Red

You hit the punching bag. It bursts, sending punching-bag-filling spraying all over the room! You know that that would happen! It always happens when you hit a punching bag! Your wife gets really angry and tells you that we don't have enough money to be getting new punching bags all the time, but women hate it when you listen to what they say! The Internet told you that!

The doorbell rings. You tear the door off its hinges instead of opening it, just to show it who's boss. Standing on your porch is a man in black. He wears a black cloak, and his face is hidden by a black hood. He raises a weapon towards you.

This looks like one of the approximately 100% of problems that can be solved by BRUTE STRENGTH! You lunge at the man, but despite your super-speed, he steps out of the way easily, even gracefully, as if he had known you were going to do that all along. He squeezes the trigger. You jump out of the way, but it turns out to be more *into* the way, as he has shot exactly where you were jumping into. Something seems very odd about this. Your last conscious thought is that you wish you had enough BRUTE STRENGTH to figure out what is going on.

Pink

You come home from work to a living room full of punching-bag-parts. Your husband isn't home. You figure he knew you were going to chew him out for destroying another punching bag, and decided to make himself scarce. That lasts right up until you go into the kitchen and see a man dressed all in black, sitting at the table, as if he was expecting you.

You panic, then reach in to touch him. If he's an axe murderer or something, you'll seduce him, get him wrapped around your little finger, then order him to jump off a cliff to prove his love for you. It's nothing you haven't done before, though you don't like to think about it too much.

Except that this man has no bare skin anywhere. His robe covers his entire body, and even his hands are gloved. You try to reach in to touch his face, but he effortlessly maneuvers away from you.

"I have your husband," he says, after you give up trying to enslave him with your magic. "He's alive and in a safe place."

"You're lying!" you answer. "He never would have surrendered to anyone! He's too alpha!"

The man nods. "I shot him with an elephant tranquilizer. He's locked up in a titanium cell underneath fifty feet of water. There's no way he can escape using BRUTE STRENGTH. If you ever want to see him again, you'll have to do what I say."

"Why? Why are you doing this to me?" you say, crying.

"I need the allegiance of some very special people," he said. "They won't listen to me just because I ask them to. But they might listen to me because *you* ask them to. I understand you are pretty special yourself. Help me get who I want, and when we are done here, I'll let you and your husband go."

There is ice in his voice. You shiver.

Grey

That night with the assassin was really scary. You swore you would never get involved in King William's business again. Why are you even considering this?

"Please?" she said, with her big puppy dog eyes.

Oh, right. Her. She's not even all that pretty. Well, pretty, but not *pretty* pretty. But somehow, when she touched you, it was like those movies where you hear a choir of angels singing in the background. You would do anything she said. You know you would.

"We need to know the layout of his palace compound," said the man in black. Was he with her? Were they dating? If they were dating, you'll kill him. It doesn't matter how creepy he is, you won't tolerate competition. But they're probably not dating. You notice how he flinches away from her, like he's afraid she might touch him.

"And it has to be me who helps?"

"I've, ah, simulated hundreds of different ways of getting access to the King. None of them hold much promise. His security is impeccable. Your special abilities are the only thing that can help us."

You sit down at your terminal. The Internet is slow; DC still doesn't have fiber optic. You've living here two years now, in a sort of retirement, ever since King William took over Russia and knocked the bottom out of the Putin-annoying business. William now controls the entire Old World, you hear, and is also Secretary-General of the United Nations and Pope of both the Catholic and the Coptic Churches. The United States is supposedly in a friendly coexistence with him, but you hear his supporters are gaining more and more power in Congress.

It only takes a few minutes' work before you have the documents you need. "He currently spends most of his time at the Rome compound," you say. "There are five different security systems. I can disable four of them. The last one is a complicated combination of electrical and mechanical that's not hooked into any computer system I'll be able to access. The only way to turn it off is from the control center, and the control center is on the *inside* of the perimeter."

The man in black nods, as if he'd been expecting that. "Come with me," he says. "We'll take care of it."

Blue

There are a hundred billion stars in the Milky Way. Each has an average of about one planet – some have many more, but a lot don't have planets at all.

If you can explore one planet every half-hour – and you can, it doesn't take too long to teleport to a planet, look around to see if there are plants and animals, and then move on to the next one – it would take you five million years to rule out life on every planet in the galaxy.

That's not practical. But, you think, life might spread. Life that originates on one planet might end up colonizing nearby planets and star systems. That means your best bet is to sample various regions of the galaxy, instead of going star by star.

That's what you've been doing. You must have seen about a hundred thousand planets so far. Some of them have beggared your imagination. Whole worlds made entirely of amethyst. Planets with dozens of colorful moons that make the night sky look like a tree full of Christmas ornaments. Planets with black inky oceans or green copper mountains.

But no life. No life anywhere.

A few years ago, you felt yourself losing touch with your humanity. You made yourself promise that every year, you'd spend a week on Earth to remind yourself of the only world you've ever seen with a population. Now it seems like an unpleasant task, an annoying imposition. But then, that was why you made yourself promise. Because you knew that future-you wouldn't do it unless they had to.

You teleport into a small Welsh hamlet. You've been away from other people so long, you might as well start small. No point going right into Times Square.

A person is standing right next to you. She reaches out her arm and touches you. You jump. How did she know you would -

"Hi," she says.

You're not a lesbian, but you can't help noticing she is the most beautiful person you've ever seen, and you would do anything for her.

"I need your help." A man dressed all in black is standing next to her.

"You should help him," the most beautiful person you've ever seen tells you, and you immediately know you will do whatever he asks.

Orange

You are in your study working on a draft version of next year's superweapon budget when you hear the door open. Four people you don't recognize step into the room. A man dressed in black. Another man wearing a grey shirt, thick glasses and is that a *pocket protector*? A woman in pink, pretty but not *pretty* pretty. Another woman in blue, who stares through you, like her mind is somewhere else. All five of your security systems have been totally silent.

You press the button to call your bodyguards, but it's not working. So you draw the gun out from under your desk and fire; you happen to be a master marksman, but the gun explodes in your face. You make a connection. A person from many years ago, who had the power to control all technology.

No time to think now. You're on your feet; good thing you happen to be a black belt in every form of martial arts ever invented. The man in grey is trying to take out a weapon; you kick him in the gut before he can get it out, and he crumples over. You go for the woman in blue, but at the last second she teleports to the other side of the room. This *isn't fair*.

You are about to go after the woman in pink, but something in her step, something in the position of the others makes you think they *want* you to attack her. You happen to be a master at reading microexpressions, so this is clear as day to you; you go after the man in black instead. He deftly sidesteps each of your attacks, almost as if he knows what you are going to do before you do it.

The woman in blue teleports behind you and kicks you in the back, hard. You fall over, and the woman in pink grabs

your hand.

She is very, very beautiful. How did you miss that before? You feel a gush of horror that you almost punched such a beautiful face.

"We need your help," she says.

You are too lovestruck to say anything.

"The pills," said the man in black. "Can you make them?"

"No," you say, truthfully. "Of course I tried. But I wouldn't even know where to begin creating magic like that."

"And you've mastered all human jobs and activities," said the man in black. "Which means the pills weren't created by any human."

"But there aren't any aliens," said the woman in blue. "Not in this galaxy, at least. I've spent years looking. It's totally dead."

"It's just as I thought," said the man in black. He turns to you. "You're the Pope now, right? Come with us. We're going to need you to get a guy in northern Italy to give us something very important."

Yellow

It is spring, now. Your favorite time in the forest. The snow has melted, the wildflowers have started to bloom, and the bears are coming out of hibernation. You're walking down to the river when someone leaps out from behind a tree and touches you. You scream, then suddenly notice how beautiful she is.

Four other people shuffle out from behind the trees. You think one of them might be King William, the new world emperor, although that doesn't really make sense.

"You're probably wondering why I've called all of you together today..." said the man in black. You're not actually wondering that, at least not in quite those terms, but the woman in pink seems be listening intently so you do the same in the hopes of impressing her.

"Somehow – and none of us can remember exactly how – each of us took a pill that gave us special powers. Mine was to see the future. I saw to the end of time, and received a message from the last people in the universe. They charged me with the task of finding the people who created these pills and asking them how entropy might be reversed.

But I couldn't do it alone. I knew there were seven other people who had taken pills. One of us – Green – is dead. Another – Red – had nothing to contribute. The rest of us are here. With the help of Pink, Blue, and Gray, we've enlisted the help of Orange and his worldwide organization. Now we're ready for the final stage of the plan. Yellow, you can read anybody's mind from a picture, right?"

Yellow nods. "But it has to be a real photograph. I can't just draw a stick figure and say it's the President and read his mind. I tried that."

Black is unfazed. "With the help of Orange, who among his many other accomplishments is the current Pope, I have

obtained the Shroud of Turin. A perfect photographic representation of Jesus Christ, created by some unknown technology in the first century. And Jesus, I am told, is an incarnation of God."

"As the current Pope, I suppose I would have to agree with that assessment," says Orange. "Though as the current UN Secretary General, I am disturbed by your fanatical religious literalism."

"Orange can do anything that humans can do, and says he can't make the pills. Blue has searched the whole galaxy, and says there aren't any aliens. That leaves only one suspect. God must have made these pills, which means He must know how to do it. If we can read His mind, we can steal his secrets."

"As Pope," says Orange, "I have to condemn this in the strongest possible terms. But as Lucasian Professor of Mathematics at Cambridge, I have to admit I'm intrigued by this opportunity to expand our knowledge."

Black ignores him. "Yellow, will you do the honors?"

You want no part in this. "This is insane. Every time I read someone's mind I regret it. Even if it's a little kid or a bear or something. It's too much for me. I can't deal with all of their guilt and sorrow and broken dreams and everything. There is *no way* I am touching the mind of God Himself."

"Pleeeeease?" asks Pink, with big puppy dog eyes.

"Um," you say.

"Don't you know how this will go, anyway?" asks Blue. "Why don't you just tell her what happens?"

"Um," said Black. "This is actually the one thing I haven't been able to see. I guess contact with God is inherently unpredictable, or something."

"I have *such* a bad feeling about this," you say.

"Pweeeeeeeease?" says Pink. She actually says pweeeeeeeease.

You sigh, take the shroud, and stare into the eyes of Weird Photographic Negative Jesus.

Black

It is the year 963,445,028,777,216 AD, and here you are in a space station orbiting the Galactic Core.

After handing Yellow the Shroud of Turin, the next thing you remember is waking up in a hospital bed. The doctor tells you that you'd been in a coma for the past forty one years.

Apparently Yellow went totally berserk after reading God's mind. You don't know the details and you don't want to, but she immediately lashed out and used her superpowers to turn off the minds of everybody within radius, including both you and herself. You all went comatose, and probably would have starved to death in the middle of the forest if Orange's supporters hadn't launched a worldwide manhunt for him. They took his body and the bodies of his friends back to Rome, where they were given the best possible medical care while a steward ruled over his empire.

After forty-one years of that, Yellow had a heart attack and died, breaking the spell and freeing the rest of you. Except

Blue and Grey. They'd died as well. It was just you, Orange, and Pink now.

Oh, and Red. You'd hired a friend to watch over him in his titanium jail cell, and once it became clear you were never coming back, he'd had mercy and released the guy. Red had since made a meager living selling the world's worst body-building videos, which were so bad they had gained a sort of ironic popularity. You tracked him down, and when Pink saw him for the first time in over forty years, she ran and embraced him. He hugged her back. It took them a few hours of fawning over each other before she realized that nothing had happened when she touched him a second time. Something something true love something the power was within you the whole time?

But you had bigger fish to fry. The stewards of Orange's empire weren't too happy about their figurehead monarch suddenly rising from the dead, and for a while his position was precarious. He asked you to be his advisor, and you accepted. With your help, he was able to retake his throne. His first act was to fund research into the immortality serum you had heard about, which was discovered right on schedule in 2060.

The years went by. Orange's empire started colonizing new worlds, then new galaxies, until thousands of years later it changed its name to the Virgo Superconfederation. New people were born. New technologies were invented. New frontiers were conquered. Until finally, the stars started going out one by one.

Faced with the impending heat death, Orange elected to concentrate all his remaining resources here, on a single station in the center of the galaxy, which would wait out the final doom as long as possible. For billions of years, it burned through its fuel stockpile, until the final doom crept closer and closer.

And then a miracle occurred.

EPILOGUE

Red

This space station is AWESOME! There are lasers and holodecks and lots of HOT PUSSY! And all you have to do is turn a giant turbine for a couple of hours a day.

One of the eggheads in white coats tried to explain it to you once. He said that your BRUTE STRENGTH was some kind of scientific impossibility, because you didn't eat or drink any more than anyone else, and you didn't breathe in any more oxygen than anyone else, and you were actually kind of small and scrawny, but you were still strong enough and fast enough to turn a giant turbine thousands of times per minute.

He rambled on and on about thermodynamics. Said that every other process in the universe used at most as much energy as you put into it, but that your strength seemed almost limitless regardless of how much energy you took in as food. That made you special, somehow. It made you a "novel power source" that could operate "independently of external negentropy". You weren't sure what any of that meant, and honestly the scientist seemed sort of like a BETA CUCKOLD ORBITER to you. But whatever was going on, they'd promised you that if you turned this turbine every day, you could have all the HOT PUSSY you wanted and be SUPER ALPHA.

You'd even met the head honcho once, a guy named King William. He told you that some of the energy you produced was going to power the station, but that the rest was going into storage. That over billions and billions of years, they would accumulate more and more stored negentropy, until it was enough to restart the universe. That it would be a cycle - a newborn universe lasting a few billion years, collapsing into a dark period when new negentropy had to be accumulated, followed by another universe again.

It all sounded way above your head. But one thing stuck with you. As he was leaving, the King remarked that it was ironic that when the black hole harvesters and wormholes and tachyon capacitors had all failed, it was a random really strong guy who had saved them.

You had always known, deep down, that BRUTE STRENGTH was what was really important. And here, at the end of all things, it is deeply gratifying to finally be proven right.

SSC Journal Club: AI Timelines

Posted on June 8, 2017 by Scott Alexander



I

A few years ago, Muller and Bostrom et al surveyed AI researchers to assess their opinion on AI progress and superintelligence. Since then, deep learning took off, AlphaGo beat human Go champions, and the field has generally progressed. I've been waiting for a new survey for a while, and now we have one.

Grace et al ([New Scientist article](#), [paper](#), see also the post on the author's blog [AI Impacts](#)) surveyed 1634 experts at major AI conferences and received 352 responses. Unlike Bostrom's survey, this didn't oversample experts at weird futurist conferences and seems to be a pretty good cross-section of mainstream opinion in the field. What did they think?

Well, a lot of different things.

The headline result: the researchers asked experts for their probabilities that we would get AI that was "able to accomplish every task better and more cheaply than human workers". The experts thought on average there was a 50% chance of this happening by 2062 – and a 10% chance of it happening by 2026!

But on its own this is a bit misleading. They also asked by what year "for any occupation, machines could be built to carry out the task better and more cheaply than human workers". The experts thought on average that there was a 50% chance of this happening by 2139, and a 20% chance of it happening by 2037.

As the authors point out, these two questions are basically the same – they were put in just to test if there was any framing effect. The framing effect was apparently strong enough to shift the median date of strong human-level AI from 2062 to 2139. This makes it hard to argue AI experts actually have a strong opinion on this.

Also, these averages are deceptive. Several experts thought there was basically a 100% chance of strong AI by 2035; others thought there was only a 20% chance or less by 2100. This is less "AI experts have spoken and it will happen in 2062" and more "AI experts have spoken, and everything they say contradicts each other and quite often themselves".

This *does* convey more than zero information. It conveys the information that AI researchers are *really unsure*. I can't tell you how many people I've heard say "there's no serious AI researcher who thinks there's any chance of human-level intelligence before 2050". Well actually, there are a few dozen conference-paper-presenting experts who think there's a *one hundred* percent chance of human-level AI before that year. I don't know what drugs they're on, but they exist. The moral of the story is: be less certain about this kind of thing.

II

The next thing we can take from this paper is a timeline of what will happen when. The authors give a bunch of different tasks, jobs, and milestones, and ask the researchers when AI will be able to complete them. Average answers

range from nearly fifty years off (for machines being able to do original high-level mathematical research) to only three years away (for machines achieving the venerable accomplishment of being able to outperform humans at *Angry Birds*). Along the way they'll beat humans at poker (four years), writing high school essays (ten years), be able to outrun humans in a 5K foot race (12 years), and write a New York Times bestseller (26 years). What do these AI researchers think is the hardest and most quintessentially human of the tasks listed, the one robots will have the most trouble doing because of its Olympian intellectual requirements? That's right – AI research (80 years).

I make fun of this, but it's actually interesting to think about. Might the AI researchers have put their own job last not because of an inflated sense of their own importance, but because they engage with it every day in Near Mode? That is, because they imagine writing a New York Times bestseller as "something something pen paper be good with words okay done" whereas they understand the complexity of AI research and how excruciatingly hard it would be to automate away every piece of what they do?

Also, since they rated AI research (80 years) as the hardest of all occupations, what do they mean when they say that "full automation of all human jobs" is 125 years away? Some other job not on the list that will take 40 years longer than AI research? Or just a combination of framing effects and not understanding the question?

(it's also unclear to what extent they believe that automating AI research will lead to a feedback loop and subsequent hard takeoff to superintelligence. This kind of theory would fit with it being the last job to be automated, but not with it taking another forty years before an unspecified age of full automation.)

III

The last part is the most interesting for me: what do AI researchers believe about risk from superintelligence?

This is very different from the earlier questions about timelines. It's possible to believe that AI will come very soon but be perfectly safe. And it's possible to believe that AI is a long time away but we really need to start preparing now, or else. A lot of popular accounts collapse these two things together, "oh, you're worried about AI, but that's dumb because there's no way it's going to happen anytime soon", but past research has shown that short timelines and high risk assessment are only modestly correlated. This survey asked about both separately.

There were a couple of different questions trying to get at this, but it looks like the most direct one was "does Stuart Russell's argument for why highly advanced AI might pose a risk, point at an important problem?". You can see the exact version of his argument quoted in the survey [on the AI Impacts page](#), but it's basically the standard Bostrom/Yudkowsky argument for why AIs may end up with extreme values contrary to our own, framed in a very normal-sounding and non-threatening way. According to the experts, this was:

No, not a real problem	11%
No, not an important problem	19%
Yes, a moderately important problem	31%
Yes, an important problem	34%
Yes, among the most important problems in the field	5%

70% of AI experts agree with the basic argument that there's a risk from poorly-goal-aligned AI. But very few believe it's among "the most important problems in the field". This is pretty surprising; if there's a good chance AI could be hostile to humans, shouldn't that automatically be pretty high on the priority list?

The next question might help explain this: “Value of working on this problem now, compared to other problems in the field?”

Much less valuable	22%
Less valuable	41%
As valuable as other problems	28%
More valuable	7%
Much more valuable	1.4%

So charitably, the answer to this question was coloring the answer to the previous one: AI researchers believe it’s plausible that there could be major problems with machine goal alignment, they just don’t think that there’s too much point in working on it now.

One more question here: “Chance intelligence explosion argument is broadly correct?”

Quite likely (81-100% chance)	12%
Likely (61-80% chance)	17%
About even (41-60% chance)	21%
Unlikely (21-40% chance)	24%
Quite unlikely (0-20% chance)	26%

Splitting the 41-60% bin in two, we might estimate that about 40% of AI researchers think the hypothesis is more likely than not.

Take the big picture here, and I worry there’s sort of a discrepancy.

50% of experts think there’s at least a ten percent chance of above-human-level AI coming within the next ten years.

And 40% of experts think that there’s a better-than-even chance that, once we get above-human level AI, it will “explode” to suddenly become vastly more intelligent than humans.

And 70% of experts think that Stuart Russell makes a pretty good point when he says that without a lot of research into AI goal alignment, AIs will probably have their goals so misaligned with humans that they could become dangerous and hostile.

I don’t have the raw individual-level data, so I can’t prove that these aren’t all anti-correlated in some perverse way that’s the opposite of the direction I would expect. But if we assume they’re not, and just naively multiply the probabilities together for a rough estimate, that suggests that about 14% of experts believe that all three of these things: that AI might be soon, superintelligent, and hostile.

Yet only a third of these – 5% – think this is “among the most important problems in the field”. Only a tenth – 1.4% – think it’s “much more valuable” than other things they could be working on.

IV

How have things changed since Muller and Bostrom's survey in 2012?

The short answer is “confusingly”. Since almost everyone agrees that AI progress in the past five years has been much faster than expected, we would expect experts to have faster timelines – ie expect AI to be closer now than they did then. But Bostrom's sample predicted human-level AI in 2040 (median) or 2081 (mean). Grace et al don't give clear means or medians, preferring some complicated statistical construct which isn't exactly similar to either of these. But their dates – 2062 by one framing, 2139 by another – at least seem potentially a little bit later.

Some of this may have to do with a subtle difference in how they asked their question:

Bostrom: “Define a high-level machine intelligence as one that can carry out most human professions as well as a typical human...”

Grace: “High-level machine intelligence is achieved when unaided machines can accomplish every task better and more cheaply than human workers.”

Bostrom wanted it equal to humans; Grace wants it better. Bostrom wanted “most professions”, Grace wants “every task”. It makes sense that experts would predict longer timescales for meeting Grace's standards.

But as we saw before, expecting AI experts to make sense might be giving them too much credit. A more likely possibility: Bostrom's sample included people from wackier subbranches of AI research, like a conference on Philosophy of AI and one on Artificial General Intelligence; Grace's sample was more mainstream. The most mainstream part of Bostrom's sample, a list of top 100 AI researchers, had an estimate a bit closer to Grace's (2050).

We can also compare the two samples on belief in an intelligence explosion. Bostrom asked how likely it was that AI went from human-level to “greatly surpassing” human level within two years. The median was 10%; the mean was 19%. The median of top AI researchers not involved in wacky conferences was 5%.

Grace asked the same question, with much the same results: a median 10% probability. I have no idea why this question – which details what an “intelligence explosion” would entail – was so much less popular than the one that used the words “intelligence explosion” (remember, 40% of experts agreed that “the intelligence explosion argument is broadly correct”). Maybe researchers believe it's a logically sound argument and worth considering but in the end it's not going to happen – or maybe they don't actually know what “intelligence explosion” means.

Finally, Bostrom and Grace both asked experts' predictions for whether the final impact of AI would be good or bad. Bostrom's full sample (top 100 subgroup in parentheses) was:

Extremely good	24%	(20)
On balance good	28%	(40)
More or less neutral	17%	(19)
On balance bad	3%	(3)
Extremely bad – existential catastrophe	18%	(8)

Grace's results for the same question:

Extremely good	20%
On balance good	25%

More or less neutral	40%
On balance bad	10%
Extremely bad – human extinction	5%

Grace's data looks pretty much the same as the TOP100 subset of Bostrom's data, which makes sense since both are prestigious non-wacky AI researchers.

V

A final question: "How much should society prioritize AI safety research"?

Much less	5%
Less	6%
About the same	41%
More	35%
Much more	12%

People who say that real AI researchers don't believe in safety research are now just empirically wrong. I can't yet say that most of them want more such research – it's only 47% on this survey. But next survey AI will be a little bit more advanced, people will have thought it over a little bit more, and maybe we'll break the 50% mark.

But we're not there yet.

I think a good summary of this paper would be that large-minorities-to-small-majorities of AI experts agree with the arguments around AI risk and think they're worth investigating further. But only a very small minority of experts consider it an emergency or think it's really important right now.

You could tell an optimistic story here – "experts agree that things will probably be okay, everyone can calm down".

You can also tell a more pessimistic story. Experts agree with a lot of the claims and arguments that suggest reason for concern. It's just that, having granted them, they're not *actually* concerned.

This seems like a pretty common problem in philosophy. "Do you believe it's more important that poor people have basic necessities of life than that you have lots of luxury goods?" "Yeah" "And do you believe that the money you're currently spending on luxury goods right now could instead be spent on charity that would help poor people get life necessities?" "Yeah." "Then shouldn't you stop buying luxury goods and instead give all your extra money beyond what you need to live to charity?" "Hey, what? Nobody does that! That would be a lot of work and make me look really weird!"

How many of the experts in this survey are victims of the same problem? "Do you believe powerful AI is coming soon?" "Yeah." "Do you believe it could be really dangerous?" "Yeah." "Then shouldn't you worry about this?" "Hey, what? Nobody does that! That would be a lot of work and make me look really weird!"

I don't know. But I'm encouraged to see people are even taking the arguments seriously. And I'm encouraged that researchers are finally giving us good data on this. Thanks to the authors of this study for being so diligent, helpful, intelligent, wonderful, and (of course) sexy.

(I might have forgotten to mention that the lead author is my girlfriend. But that's not biasing my praise above in any way.)

Asches to Asches

Posted on June 3, 2014 by Scott Alexander



Content note: fictional story contains gaslighting-type elements. May induce Cartesian skepticism.

*
*
*

You wake up in one of those pod things like in *The Matrix*. There's a woman standing in front of you, wearing a lab coat, holding a clipboard.

"Hi," she says. "This is the real world. You used to live here. We erased your memories and stuck you in a simulated world for a while, like in *The Matrix*. It was part of a great experiment."

"What?" you shout. "My whole life, a lie? How dare you deceive me as part of some grand 'experiment' I never consented to?"

"Oh," said the woman, "actually, you did consent, in exchange for extra credit in your undergraduate psychology course." She hands you the clipboard. There is a consent form with your name on it, in your handwriting.

You give her a sheepish look. "What was the experiment?"

"You know families?" asks the woman.

"Of course," you say.

"Yeah," says the woman. "Not really a thing. Like, if you think about it, it doesn't make any sense. Why would you care more for your genetic siblings and cousins and whoever than for your friends and people who are genuinely close to you? That's like racism – but even worse, at least racists identify with a group of millions of people instead of a group of half a dozen. Why should parents have to raise children whom they might not even like, who might have been a total accident? Why should people, motivated by guilt, make herculean efforts to 'keep in touch' with some nephew or cousin whom they clearly would be perfectly happy to ignore entirely?"

"Uh," you say, "not really in the mood for philosophy. Families have been around forever and they aren't going anywhere, who cares?"

"Actually," says the woman, "in the real world, no one believes in family. There's no such thing. Children are taken at birth from their parents and given to people who contract to raise them in exchange for a fixed percent of their future earnings."

"That's monstrous!" you say. "When did this happen? Weren't there protests?"

"It's always been this way," says the woman. "There's *never* been such a thing as the family. Listen. You were part of a study a lot like the [Asch Conformity Experiment](#). Our goal was to see if people, raised in a society where everyone believed X and everything revolved around X, would even be *capable* of questioning X or noticing it was stupid. We tried to come up with the stupidest possible belief, something no one in the real world had ever believed or ever

seemed likely to, to make sure that we were isolating the effect of conformity and not of there being a legitimate argument for something. So we chose this idea of ‘family’. There are racists in our world, we’re not perfect, but as far as I know none of them has *ever* made the claim that you should devote extra resources to the people genetically closest to you. That’s like a *reductio ad absurdum* of racism. So we got a grad student to simulate a world where this bizarre idea was the unquestioned status quo, and stuck twenty bright undergraduates in it to see if they would conform, or question the premise.”

“Of course we won’t question the premise, the premise is...”

“Sorry to cut you off, but I thought you should know that every single one of the other nineteen subjects, upon reaching the age where the brain they were instantiated in was capable of abstract reason, immediately determined that the family structure made no sense. One of them actually deduced that she was in a psychology experiment, because there was no other explanation for why everyone believed such a bizarre premise. The other eighteen just assumed that sometimes objectively unjustifiable ideas caught on, the same way that everyone in the antebellum American South thought slavery was perfectly natural and only a few abolitionists were able to see through it. Our conformity experiment *failed*. You were actually the only one to fall for it, hook line and sinker.”

“How could I be the only one?”

“We don’t know. Your test scores show you’re of just-above-average intelligence, so it’s not that you’re stupid. But we did give all participants a personality test that showed you have very high extraversion. The conclusion of our paper is going to be that very extraverted participants adopt group consensus without thinking and can be led to believe anything, even something as ridiculous as ‘family’.”

“I guess... when you put it like that it is kind of silly. Like, my parents were never that nice to me, but I kept loving them anyway, liking them even more than other people who treated me a lot better – and god, I even gave my mother a “WORLD’S #1 MOM” mug for Mother’s Day. That doesn’t even make sense! I... but what about the evolutionary explanation? Doesn’t evolution say we have genetic imperatives to love and support our family, whether they are worthy of it or not?”

“You can make a just-so story for *anything* using evolutionary psychology. Someone as smart as you should know better than to take them seriously.”

“But then, what *is* evolution? How did animals reproduce before the proper economic incentives were designed? Where did...”

“Tell you what. Let’s hook you up to the remnemonizer to give you your real memories back. That should answer a lot of your questions.”

A machine hovering over you starts to glow purple. “This shouldn’t hurt you a bit...”

>discontinuity<

You wake up in one of those pod things like in *The Matrix*. There’s a woman standing in front of you, wearing a lab coat, holding a clipboard.

“Hi,” she said. “There’s no such thing as virtual reality. I hypnotized you to forget all your memories from the past day and to become very confused. Then I put you in an old prop from *The Matrix* I bought off of eBay and fed you that whole story.”

"What?" you shout. "You can't just go hypnotizing and lying to people without their consent!"

"Oh," said the woman, "actually, you did consent, in exchange for extra credit in your undergraduate psychology course." She hands you the clipboard. There is a consent form with your name on it, in your handwriting. "That part was true."

You give her a sheepish look. "Why would you do such a thing?"

"Well," said the woman. "You know the Asch Conformity Experiment? I was really interested in whether you could get people to abandon some of their most fundamental beliefs, just by telling them other people believed differently. But I couldn't think of a way to test it. I mean, part of a belief being fundamental is that you already *know* everyone else believes it. There's no way I could convince subjects that the whole world was against something as obvious as 'the family' when they already know how things stand."

"So I dreamt up the weird 'virtual reality' story. I figured I would convince subjects that the real world was a lie, and that in some 'super-real' world supposedly *everybody knew* that the family was stupid, that it wasn't even an idea *worth considering*. I wanted to know how many people would give up something they've believed in for their entire life, just because they're told that 'nobody else thinks so'."

"Oh," I said. "Interesting. So even our most cherished beliefs are more fragile than we think."

"Not *really*," said the woman. "Of twenty subjects, you were the only person I got to feel any doubt, or to express any kind of anti-family sentiment."

"Frick," you say. "I feel like an idiot now. What if my mother finds out? She'll think it's her fault or something. God, she'll think I don't love her. People are going to be talking about this one *forever*."

"Don't worry," says the woman. "We'll keep you anonymized in the final data. Anyway, let's get you your memories back so you can leave and be on your way."

"You can restore my memories?" you say.

"Of course. We hypnotized you to forget the last day's events until you heard a trigger word. And that trigger is..."

>**discontinuity**<

You wake up in one of those pod things like in *The Matrix*. There's a woman standing in front of you, wearing a lab coat, holding a clipboard.

"Hi," she says. "Hypnosis is a pseudoscience and doesn't work. It was the virtual reality one, all along."

"Wut," you say.

"I mean, the first story was true. All of your memories of living with your family and so on are fake memories from a virtual world, like in *The Matrix*. The concept of 'family' really is totally ridiculous and no one in the real world believes it. All the stuff you heard first was true. The stuff about hypnosis and getting a prop from *The Matrix* off eBay was false."

"But... why?"

"We wanted to see exactly how far we could push you. You're our star subject, the only one whom we were able to induce this bizarre conformity effect in. We didn't know whether it was because you were just very very suggestible, or whether because you had never seriously considered the idea that 'family' might be insane. So we decided to do a sort of... crossover design, if you will. We took you here and debriefed you on the experiment. Then after we had told you how the world really worked, given you all the mental tools you needed to dismiss the family once and for all, even gotten you to admit we were right – we wanted to see what would happen if we sent you back. Would you hold on to your revelation and boldly deny your old society's weird prejudices? Or would you switch sides again and start acting like family made sense the second you were in a pro-family environment?"

"And I did the second one."

"Yes," says the woman. "As a psychologist, I'm supposed to remain neutral and non-judgmental. But you've got to admit, you're pretty dumb."

"Is there an experimental ethics committee I could talk to here?"

"Sorry. Experimental ethics is another one of those obviously ridiculous concepts we planted in your simulation to see if you would notice. Seriously, to believe that the progress of science should be held back by the prejudices of self-righteous fools? That's almost as weird as thinking you have a... what was the word we used... 'sister'."

"Okay, look, I realize I may have gone a little overboard helping my sister, but the experimental ethics thing seems important. Like, what's going to happen to me now?"

"Nothing's going to happen. We'll keep all your data perfectly anonymous, restore your memories, and you can be on your way."

"Um," you say. "Given past history, I'm... actually not sure I want my memories restored." You glare at the remnemonizer hovering above you. "Why don't I just..."

The woman's eyes narrow. "I'm sorry," she says. "I can't let you do that."

The machine starts to glow.

>**discontinuity**<

You wake up in one of those pod things like in *The Matrix*. There's a woman standing in front of you, wearing a lab coat, holding a clipboard.

By your count, this has happened three hundred forty six times before.

There seem to be two different scenarios. In one, the woman tells you that families exist, and have always existed. She says she has used hypnosis to make you believe in the other scenario, the one with the other woman. She asks you your feelings about families and you tell her.

Sometimes she lets you go. You go home to your mother and father, you spend some time with your sister. Sometimes you tell them what has happened. Other times you don't. You cherish your time with them, while also second-guessing everything you do. *Why* are you cherishing your time with them? Your father, who goes out drinking every night, and who has cheated on your mother more times than you can count. Your mother, who was never there for you when you needed her most. And your sister, who has been good to you, but no better than millions of other women would be, in her position. Are they a real family? Or have they been put there as a symbol of something ridiculous, impossible,

something that has never existed?

It doesn't much matter. Maybe you spend one night with them. Maybe ten. But within a month, you are always waking up in one of those pod things like in *The Matrix*.

In the second scenario, the woman tells you there are no families, never have been. She says she has used virtual reality to make you believe in the other scenario, the one with the other woman. She asks you your feelings about families and you tell her.

Sometimes she lets you go. You go to a building made of bioplastic, where you live with a carefully chosen set of friends and romantic partners. They assure you that this is how everyone lives. Occasionally, an old and very wealthy-looking man checks in with you by videophone. He reminds you that he has invested a lot of money in your upbringing, and if there's any way he can help you, anything he can do to increase your future earnings potential, you should let him know. Sometimes you talk to him, and he tells you strange proverbs and unlikely business advice.

It doesn't matter. Maybe you spend one night in your bioplastic dwelling. Maybe ten. But within a month, you are always waking up in one of those pod things like in *The Matrix*.

"Look," you tell the woman. "I'm tired of this. I know you're not bound by any kind of experimental ethics committee. But please, for the love of God, have some mercy."

"God?" asks the woman. "What does that word mean? I've never... oh right, we used *that* as our intervention in the prototype experiment. We decided 'family' made a better test idea, but Todd must have forgotten to reset the simulator."

"It's been three hundred forty six cycles," you tell her. "Surely you're not learning anything new."

"I'll be the judge of that," she says. "Now, tell me what you think about families."

You refuse. She sighs. Above you, the remnemonizer begins to glow purple.

>**discontinuity**<

You wake up in one of those pod things like in *The Matrix*. There's a purple, tentacled creature standing in front of you, wearing a lab coat, holding a clipboard.

"Hi," it says. "Turns out there's no such thing as humans."

You refuse to be surprised.

"There's only us, the 18-tkenna-dganna-07."

"Okay," you say. "I want answers."

"Absolutely," says the alien. "We would like to find optimal social arrangements."

"And?"

"And I cannot tell you whether we have families or not, for reasons that are to become apparent, but the idea is at least sufficiently interesting to have entered the space of hypotheses worth investigating. But we don't trust ourselves to

investigate this. It's the old Asch Conformity Problem again. If we have families, then perhaps the philosophers tasked with evaluating families will conform to our cultural norms and decide we should keep them. If we do not, perhaps the philosophers will conform and decide we should continue not to. So we determined a procedure that would create an entity capable of fairly evaluating the question of families, free from conformity bias."

"And that's what you did to me."

"Yes. Only by exposing you to the true immensity of the decision, without allowing you to fall back on what everyone else thinks, could we be confident in your verdict. Only by allowing you to experience both how obviously right families are, when you 'know' they are correct, and how obviously wrong families are, when you 'know' they are incorrect, could we expect you to garner the wisdom to be found on both sides of the issue."

"I see," you say, and you do.

"Then, O purified one," asks the alien, "tell us of your decision."

"Well," you say. "If you have to know, I think there are about equally good points on both sides of the issue."

"Fuck," says the 18-tkenna-dganna-07.

SSC Journal Club: Dissolving The Fermi Paradox

Posted on July 3, 2018 by Scott Alexander



I'm late to posting this, but it's important enough to be worth sharing anyway: [Sandberg, Drexler, and Ord on Dissolving the Fermi Paradox](#).

(You may recognize these names: Toby Ord founded the effective altruism movement; Eric Drexler kindled interest in nanotechnology; Anders Sandberg helped pioneer the academic study of x-risk, and [wrote what might be my favorite *Unsong* fanfic](#))

The Fermi Paradox asks: given the immense number of stars in our galaxy, for even a very tiny chance of aliens per star shouldn't there be thousands of nearby alien civilizations? But any alien civilization that arose millions of years ago would have had ample time to colonize the galaxy or do something equally dramatic that would leave no doubt as to its existence. So where are they?

This is sometimes formalized as the Drake Equation: think up all the parameters you would need for an alien civilization to contact us, multiply our best estimates for all of them together, and see how many alien civilizations we predict. So for example if we think there's a 10% chance of each star having planets, a 10% chance of each planet being habitable to life, and a 10% chance of a life-habitable planet spawning an alien civilization by now, one in a thousand stars should have civilization. The actual Drake Equation is much more complicated, but most people agree that our best-guess values for most parameters suggest a vanishingly small chance of the empty galaxy we observe.

SDO's contribution is to point out this is the wrong way to think about it. [Sniffnoy's comment on the subreddit](#) helped me understand exactly what was going on, which I think is something like this:

Imagine we knew God flipped a coin. If it came up heads, He made 10 billion alien civilization. If it came up tails, He made none besides Earth. Using our one parameter Drake Equation, we determine that *on average* there should be 5 billion alien civilizations. Since we see zero, that's quite the paradox, isn't it?

No. In this case the mean is meaningless. It's not at all surprising that we see zero alien civilizations, it just means the coin must have landed tails.

SDO say that relying on the Drake Equation is the same kind of error. We're not interested in the *average* number of alien civilizations, we're interested in the distribution of probability over number of alien civilizations. In particular, what is the probability of few-to-none?

SDO solve this with a "synthetic point estimate" model, where they choose random points from the distribution of possible estimates suggested by the research community, run the simulation a bunch of times, and see how often it returns different values.

According to their calculations, a standard Drake Equation multiplying our best estimates for every parameter together yields a probability of less than one in a million billion billion billion that we're alone in our galaxy – making such an observation pretty paradoxical. SDO's own method, taking account parameter uncertainty into account, yields a probability of one in three.

They try their hand at doing a Drake calculation of their own, using their preferred values, and find:

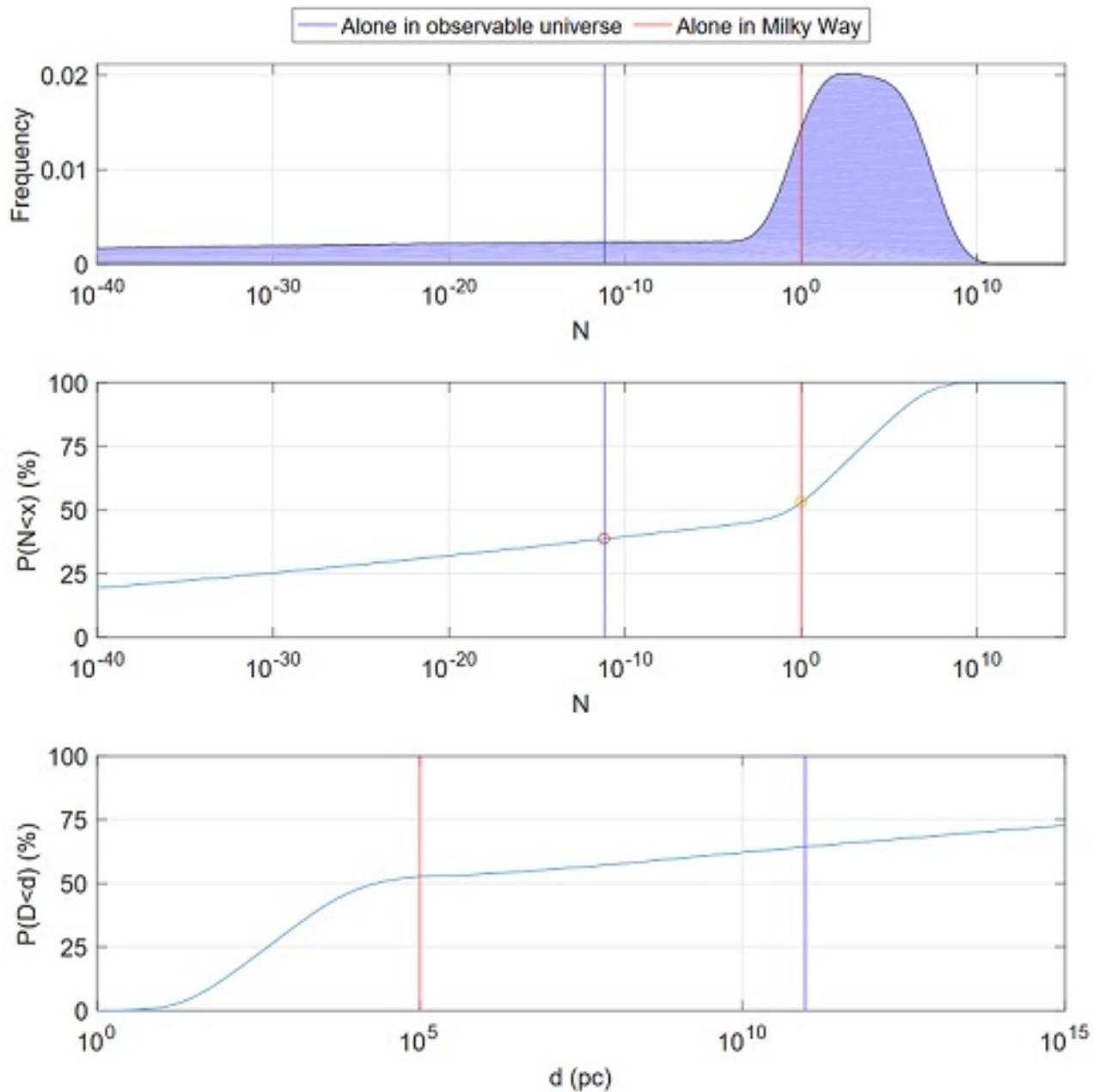


Figure 2: (A) A probability density function for N generated by Monte Carlo simulation based on the authors' best estimates of our current uncertainty for each parameter. (B) The corresponding cumulative density function. (C) A cumulative density function for the distance to the nearest detectable civilization.

N is the average number of civilizations per galaxy

If this is right – and we can debate exact parameter values forever, but it's hard to argue with their point-estimate-vs-distribution-logic – then there's no Fermi Paradox. It's done, solved, kaput. Their title, "Dissolving The Fermi Paradox", is a strong claim, but as far as I can tell they totally deserve it.

"Why didn't anyone think of this before?" is the question I am only slightly embarrassed to ask given that I didn't think of it before. I don't know. Maybe people thought of it before, but didn't publish it, or published it somewhere I don't know about? Maybe people intuitively figured out what was up (one of the parameters of the Drake Equation must be much lower than our estimate) but stopped there and didn't bother explaining the formal probability argument. Maybe [nobody took the Drake Equation seriously anyway](#), and it's just used as a starting point to discuss the probability of life forming?

But any explanation of the "oh, everyone knew this in some sense already" sort has to deal with that a lot of very smart and well-credentialled experts treated the Fermi Paradox very seriously and came up with all sorts of weird explanations. There's no need for sci-fi theories any more (though you should still read [the Dark Forest trilogy](#)). It's just that there aren't very many aliens. I think my past speculations on this, though very incomplete and much inferior to the recent paper, [come out pretty well](#) here.

(some more discussion [here](#) on Less Wrong)

One other highlight hidden in the supplement: in the midst of a long discussion on the various ways intelligent life can fail to form, starting on page 6 the authors speculate on "alternative genetic systems". If a planet gets life with a slightly different way of encoding genes than our own, it might be too unstable to allow complex life, or too stable to allow a reasonable rate of mutation by natural selection. It may be that abiogenesis can only create very weak genetic codes, and life needs to go through several "genetic-genetic transitions" before it can reach anything capable of complex evolution. If this is path-dependent – ie there are branches that are local improvements but close off access to other better genetic systems – this could permanently arrest the development of life, or freeze it at an evolutionary rate so low that the history of the universe so far is too short a time to see complex organisms.

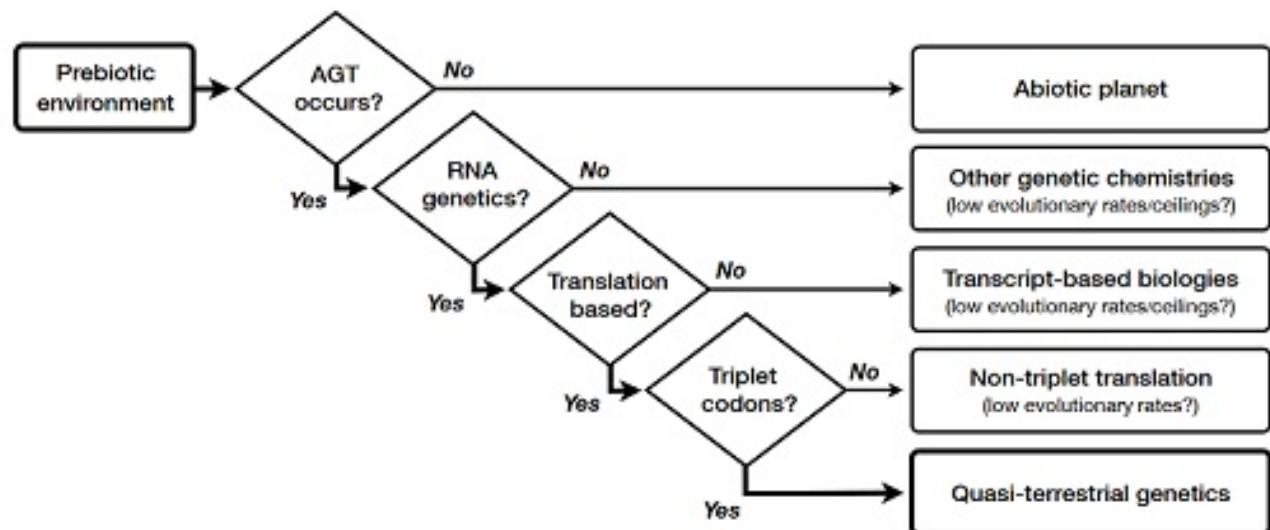


Figure 1. A schematic diagram of potential AGT and GGT branches with aggregated states and transitions. At each branch point, outcome ratios are weakly constrained, with LU values arguably > 100.

I don't claim to understand all of this, but the parts I do understand are fascinating and could easily be their own paper.

Book Review: Reframing Superintelligence

Posted on August 27, 2019 by Scott Alexander



I

Ten years ago, everyone was talking about superintelligence, the singularity, the robot apocalypse. What happened?

I think the main answer is: the field matured. Why isn't everyone talking about nuclear security, biodefense, or counterterrorism? Because there are already competent institutions working on those problems, and people who are worried about them don't feel the need to take their case directly to the public. The past ten years have seen AI goal alignment reach that level of maturity too. There are all sorts of new research labs, think tanks, and companies working on it – the [Center For Human-Compatible AI](#) at UC Berkeley, [OpenAI](#), [Ought](#), the [Center For The Governance Of AI](#) at Oxford, the [Leverhulme Center For The Future Of Intelligence](#) at Cambridge, etc. Like every field, it could still use more funding and talent. But it's at a point where academic respectability trades off against public awareness at a rate where webzine articles saying CARE ABOUT THIS OR YOU WILL DEFINITELY DIE are less helpful.

One unhappy consequence of this happy state of affairs is that it's harder to keep up with the field. In 2014, Nick Bostrom wrote *Superintelligence: Paths, Dangers, Strategies*, giving a readable overview of what everyone was thinking up to that point. Since then, things have been less public-facing, less readable, and more likely to be published in dense papers with a lot of mathematical notation. They've also been – no offense to everyone working on this – less revolutionary and less interesting.

This is one reason I was glad to come across [Reframing Superintelligence: Comprehensive AI Services As General Intelligence](#) by Eric Drexler, a researcher who works alongside Bostrom at Oxford's Future of Humanity Institute. This 200 page report is not quite as readable as *Superintelligence*; its highly-structured outline form belies the fact that all of its claims start sounding the same after a while. But it's five years more recent, and presents a very different vision of how future AI might look.

Drexler asks: what if future AI looks a lot like current AI, but better?

For example, take Google Translate. A future superintelligent Google Translate would be able to translate texts faster and better than any human translator, capturing subtleties of language beyond what even a native speaker could pick up. It might be able to understand hundreds of languages, handle complicated multilingual puns with ease, do all sorts of amazing things. But in the end, it would just be a translation app. It wouldn't want to take over the world. It wouldn't even "want" to become better at translating than it was already. It would just translate stuff really well.

The future could contain a vast ecosystem of these superintelligent services before any superintelligent agents arrive. It could have media services that can write books or generate movies to fit your personal tastes. It could have invention services that can design faster cars, safer rockets, and environmentally friendly power plants. It could have strategy services that can run presidential campaigns, steer Fortune 500 companies, and advise governments. All of them would be far more effective than any human at performing their given task. But you couldn't ask the presidential-campaign-running service to design a rocket any more than you could ask Photoshop to run a spreadsheet.

In this future, our AI technology would have taken the same path as our physical technology. The human body can run fast, lift weights, and fight off enemies. But the automobile, crane, and gun are three different machines. Evolution had to cram running-ability, lifting-ability, and fighting-ability into the same body, but humans had more options and were able to do better by separating them out. In the same way, evolution had to cram book-writing, technology-inventing, and strategic-planning into the same kind of intelligence – an intelligence that also has associated goals and drives. But humans don't have to do that, and we probably won't. We're not doing it today in 2019, when Google Translate and AlphaGo are two different AIs; there's no reason to write a single AI that both translates languages and plays Go. And we probably won't do it in the superintelligent future either. Any assumption that we will is based more on anthropomorphism than on a true understanding of intelligence.

These superintelligent services would be safer than general-purpose superintelligent agents. General-purpose superintelligent agents (from here on: agents) would need a human-like structure of goals and desires to operate independently in the world; Bostrom has explained ways this is likely to go wrong. AI services would just sit around algorithmically mapping inputs to outputs in a specific domain.

Superintelligent services would not self-improve. You could build an AI researching service – or, more likely, several different services to help with several different aspects of AI research – but each of them would just be good at solving certain AI research problems. It would still take human researchers to apply their insights and actually build something new. In theory you might be able to automate every single part of AI research, but it would be a weird idiosyncratic project that wouldn't be anybody's first choice.

Most important, superintelligent services could help keep the world safe from less benevolent AIs. Drexler agrees that a self-improving general purpose AI agent is possible, and assumes someone will build one eventually, if only for the lulz. He agrees this could go about the way Bostrom expects it to go, ie very badly. But he hopes that there will be a robust ecosystem of AI services active by then, giving humans superintelligent help in containing rogue AIs. Superintelligent anomaly detectors might be able to notice rogue agents causing trouble, superintelligent strategic planners might be able to develop plans for getting rid of them, and superintelligent military research AIs might be able to create weapons capable of fighting them off.

Drexler therefore does not completely dismiss Bostromian disaster scenarios, but thinks we should concentrate on the relatively mild failure modes of superintelligent AI services. These may involve normal bugs, where the AI has aberrant behaviors that don't get caught in testing and cause a plane crash or something, but not the unsolveable catastrophes of the Bostromian paradigm. Drexler is more concerned about potential misuse by human actors – either illegal use by criminals and enemy militaries, or antisocial use to create things like an infinitely-addictive super-Facebook. He doesn't devote a lot of space to these, and it looks like he hopes these can be dealt with through the usual processes, or by prosocial actors with superintelligent services on their side (thirty years from now, maybe people will say "it takes a good guy with an AI to stop a bad guy with an AI").

This segues nicely into some similar concerns that OpenAI researcher Paul Christiano has brought up. He worries that AI services will be naturally better at satisfying objective criteria than at "making the world better" in some vague sense. Tasks like "maximize clicks to this site" or "maximize profits from this corporation" are objective criteria; tasks like "provide real value to users of this site instead of just clickbait" or "have this corporation act in a socially responsible way" are vague. That means AI may asymmetrically empower some of the worst tendencies in our society without giving a corresponding power increase to normal people just trying to live enjoyable lives. In his model, one of the tasks of AI safety research is to get AIs to be as good at optimizing vague prosocial tasks as they will naturally be at optimizing the bottom line. Drexler doesn't specifically discuss this in *Reframing Superintelligence*, but it seems to fit the spirit of the kind of thing he's concerned about.

II

I'm not sure how much of the AI alignment community is thinking in a Drexlerian vs. a Bostromian way, or whether that is even a real dichotomy that a knowledgeable person would talk about. I know there are still some people who are very concerned that even programs that seem to be innocent superintelligent services will be able to self-improve, develop misaligned goals, and cause catastrophes. I got to talk to Dr. Drexler a few years ago about some of this (although I hadn't read the book at the time, didn't understand the ideas very well, and probably made a fool of myself); at the time, he said that his work was getting a mixed reception. And there are still a few issues that confuse me.

First, many tasks require general intelligence. For example, an AI operating in a domain with few past examples (eg planning defense against a nuclear attack) will not be able to use modern training paradigms. When humans work on these domains, they use something like common sense, which is presumably the sort of thing we have because we understand thousands of different domains from gardening to ballistics and this gives us a basic sense of how the world works in general. Drexler agrees that we will want AIs with domain-general knowledge that cannot be instilled by training, but he argues that this is still "a service". He agrees these tasks may require AI architectures different from any that currently exist, with relatively complete world-models, multi-domain reasoning abilities, and the ability to learn "on the fly" – but he doesn't believe those architectures will need to be agents. Is he right?

Second, is it easier to train services or agents? Suppose you want a good multi-domain reasoner that can help you navigate a complex world. One proposal is to create AIs that train themselves to excel in world simulations the same way AlphaGo trained itself to excel in simulated games of Go against itself. This sounds a little like the evolutionary process that created humans, and agent-like drives might be a natural thing to come out of this process. If agents were easier to "evolve" than services, agentic AI might arise at an earlier stage, either because designers don't see a problem with it or because they don't realize it is agentic in the relevant sense.

Third, how difficult is it to separate agency from cognition? Natural intelligences use "active sampling" strategies at levels as basic as sensory perception, deciding how to direct attention in order to best achieve their goals. At higher levels, they decide things like which books to read, whose advice to seek out, or what subdomain of the problem to evaluate first. So far AIs have managed to address even very difficult problems without doing this in an agentic way. Can this continue forever? Or will there be some point at which intelligences with this ability outperform those without it.

I think Drexler's basic insight is that Bostromian agents need to be really different from our current paradigm to do any of the things Bostrom predicts. A paperclip maximizer built on current technology would have to eat gigabytes of training data about various ways people have tried to get paperclips in the past so it can build a model that lets it predict what works. It would build the model on its actually-existing hardware (not an agent that could adapt to much better hardware or change its hardware whenever convenient). The model would have a superintelligent understanding of the principles that had guided some things to succeed or fail in the training data, but wouldn't be able to go far beyond them into completely new out-of-the-box strategies. It would then output some of those plans to a human, who would look them over and make paperclips 10% more effectively.

The very fact that this is less effective than the Bostromian agent suggests there will be pressure to build the Bostromian agent eventually (Drexler disagrees with this, but I don't understand why). But this will be a very different project from AI the way it currently exists, and if AI the way it currently exists can be extended all the way to superintelligence, that would give us a way to deal with hostile superintelligences in the future.

III

All of this seems kind of common sense to me now. This is worrying, because I didn't think of any of it when I read *Superintelligence* in 2014.

I asked readers to tell me if there was any past discussion of this. Many people brought up [Robin Hanson's arguments](#), which match the “ecosystem of many AIs” part of Drexler’s criticisms but don’t focus as much on services vs. agents. Other people brought up discussion under the heading of [Tool AI](#). Combine those two strains of thought, and you more or less have Drexler’s thesis, minus some polish. I read some of these discussions, but I think I failed to really understand them at the time. Maybe I failed to combine them, focused too much on the idea of an Oracle AI, and missed the idea of an ecosystem of services. Or maybe it all just seemed too abstract and arbitrary when I had fewer examples of real AI systems to think about.

I’ve sent this post by a couple of other people, who push back against it. They say they still think Bostrom was right on the merits and superintelligent agents are more likely than superintelligent services. Many brought up [Gwern’s essay on why tool AIs are likely to turn into agent AIs](#) and [this post by Eliezer Yudkowsky](#) on the same topic – I should probably reread these, reread Drexler’s counterarguments, and get a better understanding. For now I don’t think I have much of a conclusion either way. But I think I made a [mistake of creativity](#) in not generating or understanding Drexler’s position earlier, which makes me more concerned about how many other things I might be missing.

It Was You Who Made My Blue Eyes Blue

Posted on October 15, 2015 by Scott Alexander



Content note: suicide.

Day Zero

It all started with an ignorant white guy.

His name was Alonzo de Pinzon, and he'd been shipwrecked. We heard him yelling for help on the rocks and dragged him in, even though the storm was starting to get really bad. He said that his galleon had gone down, he'd hung on to an oar and was the only survivor. Now he was sitting in our little hunting lodge, shivering and chattering his teeth and asking us questions in the Polynesian traders' argot which was the only language we all shared.

"How big is this island? How many of you are there?"

Daho answered first. "11.8 miles from the easternmost point to the westernmost point, 3.6 miles from the northernmost to the southernmost. Total area is 14.6 square miles, total coastline is dependent on how deeply you want to go into the fractal nature of the perimeter but under some reasonable assumptions about 32 miles long. Last census said there were 906 people, but that was two years ago, so assuming the 5.1% rate of population growth continues, there should be closer to 1000 now. Everyone else is back at the village, though. The five of us were out hunting and got caught in the storm. We figured we'd stay at this old hunting lodge until it cleared up, since it's 5.5 miles back to the village and given the terrain and factoring in a delay because of the storm it would probably take at least 9.5 hours to get back."

Pinzon blinked.

"Problem?" asked Daho.

"But – " he said. "That is the sort of answer I should expect from a natural philosopher. Not from a savage."

"Savage?" Calkas hissed. "Really? We rescue you, and the first thing you do is call us savages?"

The sailor looked around, as if anxious. Finally, almost conspiratorially: "But I heard about your island! I heard you eat people!"

Calkas smiled. "Only as a deterrent. Most of the time when European explorers land somewhere, they kill all the men and enslave all the women and convert the children to Christianity. The only places that escape are the ones that get a reputation for eating said European explorers. So we arranged to give ourselves that reputation."

"And then we had to go through with it a few times in order to make the deterrent credible," added Bekka, my betrothed. "And you guys do taste really good with ketchup."

"It's a savage thing to do!" Pinzon said "And you even look like savages. You wear bones in your hair"

"Just Enuli," I said. "She's going through a Goth phase."

"My name is Morticia now," said Enuli, "and it's *not a phase!*" She did have a bone in her hair. She also had white face paint and black eyeliner.

"More roast pig?" Bekka asked Pinzon. The sailor nodded, and she re-filled his plate.

"I just don't get it," he told us. "Everyone else in this part of the world lives in thatched huts and counts 'one, two, many'. We tried to trade with the Tahitians, and they didn't understand the concept of money! It was a mess!"

Bekka rolled her eyes at me, and I smiled. Calkas was a little more tolerant. "The sacred plant of our people is called sparkroot," he said. "When we eat it, we get – more awake, I guess you could say. We try to have some every day, and it helps us keep track of things like the island size and the population, and much more."

Alonzo de Pinzon looked interested. "How come you haven't done more with your intellect? Invented galleons, like we Spaniards? Set off to colonize Tahiti or the other islands? If you are as smart as you seem, you could conquer them and take their riches."

"Maybe," said Calkas. "But that's not why the Volcano God gave us the sparkroot. He gave us sparkroot to help us comply with his complicated ritual laws."

"You need to be smart to deal with your ritual laws?"

"Oh yes. For example, the Tablets of Enku say that we must count the number of days since Enku The Lawgiver first spoke to the Volcano God, and on days whose number is a Mersenne prime we can't eat any green vegetables."

"What's a Mersenne prime?" asked the sailor.

"Exactly my point," said Calkas, smiling.

"That's not even the worst of it!" Daho added. "The Tablets say we have to bathe in the waterfall any day x such that $a^n + b^n = x^n$ where n is greater than two. We got all confused by that one for a while, until Kaluhani gorged himself on a whole week's worth of sparkroot in one night and proved that it would never apply to any day at all."

"The Volcano God's yoke is light," Calkas agreed.

"Although poor Kaluhani was vomiting for the next three days after that," Bekka reminded us, and everybody laughed remembering.

"Oh!" said Daho. "And remember that time when Uhuako was trying to tattoo everyone who didn't tattoo themselves, and he couldn't figure out whether he had to tattoo himself or not, so he ended up eating a whole sparkroot plant at once and inventing advanced set theory? That was hilarious."

Everyone except Alonzo de Pinzon giggled.

"Point is," said Calkas, "that's why the Volcano God gives us sparkroot. To follow the rituals right. Any other use is taboo. And I'm okay with that. You Europeans may have your big ships and your guns and your colonies across half the world. And you might think you're smart. But you guys couldn't follow the Volcano God's rituals right for a *day* without

your brains exploding."

Pinzon scowled. "You know what?" he said. "I don't think you're Polynesians at all. I think you must be descended from Europeans. Maybe some galleon crashed on this island centuries ago, and you're the descendants. That would explain why you're so smart."

"You know what else we've invented with our giant brains?" Bekka asked. "Not being racist."

"It's not racism!" said Pinzon. "Look, there's one more obvious reason to think you're descended from Europeans. You may have dark skin, but this is the first place I've been in all of Polynesia where I've seen even one native with blue eyes."

Bekka gasped. Calkas' eyes went wide. Daho's hands started curling into fists. Enuli started to sob.

I looked at them. They looked at me. Then, as if synchronized, we grabbed Alonzo de Pinzon and crushed his throat and held him down until he stopped breathing.

He tasted delicious with ketchup.

Day One

The next morning dawned, still grey and cold and stormy.

"So," I said when the other four had awoken. "I guess we're all still here."

I said it glumly. It wasn't that I wanted any of my friends to commit suicide. But if one of them had, the horror would have stopped there. Of course, I knew it couldn't really be over that easily. But I couldn't have admitted I knew. I couldn't even have suggested it. That would have made me as bad as the Spanish sailor.

"Wait," said Enuli. "I don't get it. Why wouldn't we still be here?"

The other four stared at her like she was mad.

"Enuli," Calkas suggested, "did you forget your sparkroot last night?"

"First of all, my name is Morticia. And - "

"Shut it. Did you forget your sparkroot?"

Finally she nodded bashfully. "I was so upset about that awful man making fun of my hair-bone," she said. "I guess it slipped my mind. I'll have some now." She took some raw sparkroot from our bag, started to crush it with the mortar and pestle. "In the meantime, tell me what's going on."

"Alonzo de Pinzon said at least one of us had blue eyes. We all know what the Tablets of Enku say. If anybody has blue eyes, and knows that they have blue eyes, they must kill themselves."

"So what? I see people with blue eyes all the time. Of course at least one of us has blue eyes."

Concerned looks from the others. I reflected for a second, the sparkroot smoothing the thoughts' paths through my brain. No, she hadn't revealed anything extra by saying that, although she would have if she had said it before the sailor had spoken, or last night before we woke up this morning. She hadn't made the problem worse. Still, it had been a slip. This was the sort of thing that made forgetting your sparkroot so dangerous. Had it been a different time, even Enuli's comment could have doomed us all.

"It's like this," I told Enuli. "Suppose there were only the two of us, and we both had blue eyes. Of course, you could see me and know that I had blue eyes. So you would know that at least one of us had blue eyes. But what you wouldn't know is that I also knew it. Because as far as you know, you might have eyes of some other color, let's say brown eyes. If you had brown eyes, and I of course don't know my own eye color, then I would still think it possible that both of us have brown eyes. So if I in fact know for sure that at least one of us has blue eyes, that means you have blue eyes. So you know at least one of us has blue eyes, but you don't know that I know it. But if Alonzo de Pinzon shows up and says that at least one of us has blue eyes, now you know that I know it."

"So?" Enuli poured the ground-up root into a cup of boiling water.

"So the Tablets say that if anyone knows their own eye color, they must commit suicide at midnight of that night. Given that I know at least one of us has blue eyes, if I see you have brown eyes, then I know my own eye color – I must be the blue-eyed one. So the next morning, when you wake up and see me not dead, you know that you don't have brown eyes. That means you must be the blue-eyed one. And that means you have to kill yourself on midnight of the following night. By similar logic, so do I."

Enuli downed her sparkroot tea, and then her eyes lit up. "Oh, of course," she said. Then "Wait! If we follow the situation to its logical conclusion, any group of n blue-eyed people who learn that at least one of them has blue eyes have to kill themselves on the nth night after learning that!"

We all nodded. Enuli's face fell.

"I don't know about the rest of you," said Daho, "but I'm not just going to sit around and wait to see if I die." There were murmurs of agreement.

I looked out at my friends. Four pairs of blue eyes stared back at me. Everybody else either saw four pairs of blue eyes or three pairs of blue eyes, depending on what color my own eyes were. Of course, I couldn't say so aloud; that would speed up the process and cost us precious time. But I knew. And they knew. And I knew they knew. And they knew I knew I knew. Although they didn't know I knew they knew I knew. I think.

Then I looked at Bekka. Her big blue eyes stared back at me. There was still hope I was going to survive this. My betrothed, on the other hand, was absolutely doomed.

"This sucks," I agreed. "We've got to come up with some kind of plan. Maybe – Enuli wasn't thinking straight yesterday. So her not committing suicide doesn't count. Can we work with that?"

"No," said Calkas. "Suppose Enuli was the only one with blue eyes, and all the rest of us had brown eyes. Then she would realize that and commit suicide tonight. If she doesn't commit suicide tonight, then we're still screwed."

"Um," said Daho. "I hate to say this, but we get rid of Enuli. There's a canoe a little ways down the beach hidden underneath the rocks. She can set off and row for Tahiti. We'll never know if she killed herself tonight or not. Remember, right now for all we know Enuli might be the only one with blue eyes. So if there's any question in our mind about whether she killed herself, we can't be sure that the rest of us aren't all brown-eyed."

We all thought about that for a moment.

"I'm not going to row to Tahiti," said Enuli. "In this storm, that would be suicide."

The rest of us glared at her.

"If you don't get off this island, then for all we know all five of us are going to have to die," I said. "You included."

"Well Ahuja, if you're so big on making sacrifice why don't *you* go to Tahiti?"

"First of all," I said, "because I'm not leaving my betrothed. Second of all, because it doesn't work for me. I knew what was going on last night. We already know that I'm not the only blue-eyed person here. And we know we know it, and know we know we know it, and so on. You're the only one who can help us."

"Yeah?" said Enuli. "Well, if two of you guys were to row to Tahiti, that would solve the problem too."

"Yes," said Daho patiently. "But then two of us would be stuck in exile. If you did it, only one of us would be stuck."

Enuli gave a wicked grin. "You know what?" she said. "I'll say it. I'm not the only blue-eyed person here. At least one of the rest of you has blue eyes."

And there it was.

"Ha. Now I'm no worse off than any of the rest of you."

"Kill her," said Bekka. "She broke the taboo." The rest of us nodded.

"So she did," said Calkas. "And if we had a court here, led by the high priest, and an executioner's blade made to exactly the right standard, kill her we would. But until those things happen, it is taboo for us to convict and kill her without trial."

Calkas' father was the high priest. He knew the law better than any of us. The five of us sat quietly and thought about it. Then he spoke again:

"But her soul may well burn in the caldera of the Volcano God forever."

Enuli started to cry.

"And," Calkas continued, "there is nevertheless a flaw in our plan. For all we know, three out of five of us have brown eyes. We cannot tell the people who have blue eyes that they have blue eyes without breaking the taboo. So we cannot force blue-eyed people in particular to sail to Tahiti. But if two of the brown-eyed people sail to Tahiti, then we do not lose any information; we know that they would not have committed suicide, because they could not have figured out their own eye color. So sailing to Tahiti won't help."

The rest of us nodded. Calkas was right.

"Let's wait until dinner tonight," I suggested. "We'll all have some more sparkroot, and maybe we'll be able to think about the problem a little more clearly."

Day Two

The sun rose behind angry storm clouds. The five of us rose with it.

"Well, I guess we're all still here," I said, turning the morning headcount into a grim tradition.

"Look," said Bekka. "The thing about sailing to Tahiti would work a lot better if we knew how many blue-eyed versus brown-eyed people were here. If we all had blue eyes, then we could be sure that the Tahiti plan would work, and some of us could be saved. If some of us had brown eyes, then we could choose a number of people to sail to Tahiti that had a good probability of catching enough of the blue-eyed ones."

"We can wish all we want," said Enuli, "but if we explicitly knew how many people had blue versus brown eyes, we'd all have to kill ourselves right now."

"What about probabilistic knowledge?" I asked. "In theory, we could construct a system that would allow us to have > 99.99% probability what color our eyes were without being sure."

"That's stupid," Enuli said, at precisely the same time Calkas said "That's brilliant!" He went on: "Look, just between the five of us, everybody else back at the village has blue eyes, right?"

We nodded. It was nerve-wracking to hear it mentioned so casually, just like that, but as far as I could tell it didn't break any taboos.

"So," said Calkas, "We know that, of the island population, at least 995 of the 1000 of us have blue eyes. Oh, and since nobody committed suicide last night, we know that at least three of the five of us have blue eyes, so that's 998 out of 1000. Just probabilistically, by Laplace's Law of Succession and the like, we can estimate a >99% chance that we ourselves have blue eyes. Nothing I'm saying is taboo. It's nothing that the priests don't know themselves. But none of them have killed themselves yet. So without revealing any information about the eye color composition of the current group, I think it's reasonable to make a first assumption that all of us have blue eyes."

"I'm really creeped out at you talking like this," said Daho. I saw goosebumps on his arms.

"I do not believe that the same Volcano God who has endowed us with reason and intellect could have intended us to forego their use," said Calkas. "Let's assume we all have blue eyes. In that case, the Tahiti plan is still on."

"Waaiiiit a second –" Bekka objected. "If probabilistic knowledge of eye color doesn't count, then no information can count. After all, there's always a chance that the delicious sailor could have been lying. So when he said at least one of us had blue eyes, all we know is that there's a high *probability* that at least one of us has blue eyes."

"Yes!" said Daho. "I've been reading this book that washed ashore from a shipwrecked galleon. Off in Europe, there is this tribe called the Jews. Their holy book says that illegitimate children should be shunned by the congregation. Their leaders thought this was unfair, but they weren't able to contradict the holy book. So instead they declared that sure, illegitimate children should be shunned, but only if they were *sure* they were really illegitimate. Then they declared that no amount of evidence would ever suffice to convince them of that. There was always a possibility that the woman had secretly had sex with her husband nine months before the birth and was simply lying about it. Or, if apparently unmarried, that she had secretly married someone. They decided that it was permissible to err on the side of caution, and from that perspective nobody was sufficiently certainly illegitimate to need shunning. We could do the same thing here."

"Yes!" I said. "That is, even if we looked at our reflection and saw our eye color directly, it might be that a deceiving

demon is altering all of our experience - ”

“No no NO,” said Calkas. “That’s not right. The Tablets of Enku say that *because* people must not know their own eye color, we are forbidden to talk about the matter. So the law strongly implies that hearing someone tell us our eye color would count as proof of that eye color. The exact probability has nothing to do with it. It’s the method by which we gain the information.”

“That’s stupid,” Bekka protested.

“That’s the law,” said Calkas.

“Let’s do the Tahiti plan, then,” I said. I gathered five stones from the floor of the lodge. Two white, three black. “White stones stay. Black stones go to Tahiti. Close your eyes and don’t look.”

Bekka, Calkas, Daho, and Enuli all took a stone from my hand. I looked at the one that was left. It was black. Then I looked around the lodge. Calkas and Enuli were smiling, white stones in their hands. Bekka and Daho, not so much. Daho whined, looked at me pleadingly.

“No,” I said. “It’s decided. The three of us will head off tonight.”

Calkas and Enuli tried to be respectful, to hide their glee and relief.

“You guys will tell our families what happened?

They nodded gravely.

We began packing our things.

*
*

The dark clouds frustrated any hope of moonlight as Bekka, Daho and I set off to the nearby cove where two canoes lay hidden beneath the overhanging rocks. The rain soaked our clothes the second we crossed the doorway. The wind lashed at our faces. We could barely hear ourselves talk. This was a *bad* storm.

“How are we going to make it to the canoes in this weather?!” Bekka shouted at me, grabbing my arm. I just squeezed her hand. Daho might have said something, might not have. I couldn’t tell. Between the mud and the rain and the darkness it took us two hours to travel less than a mile. The canoes were where we had left them a few days before. The rocks gave us brief shelter from the pelting rain.

“This is suicide!” Daho said, once we could hear each other again. “There’s no way we can make it to Tahiti in this! We won’t even be able to make it a full mile out!” Bekka nodded.

“Yes,” I said. I’d kind of known it, the whole way down to the cove, but now I was sure. “Yes. This is suicide. But we’ve got to do it. If we don’t kill ourselves tonight, then we’ve just got to go back to the lodge. And then we’ll all end up killing ourselves anyway. And Calkas and Enuli will die too.”

“No!” said Daho. “We go back, we tell them that we can’t make it to Tahiti. Then we let *them* decide if we need to commit suicide or not. And if they say yes, we draw the stones again. Four black, one white. One chance to live.”

“We already drew the stones,” I said. “Fair is fair.”

"Fair is fair?" Bekka cried. "We drew stones to go to Tahiti. We didn't draw stones to commit suicide. If the stone drawing obliged us to commit suicide, they should have said so, and then maybe we would have spent more time thinking about other options. Why do we have to die? Why can't the other ones die? Why not Enuli, with that stupid bone in her hair? I hate her so much! Ahuja, you can't just let me die like this!"

That hurt. I was willing to sacrifice my life, if that was what it took. But Bekka was right. To just toss ourselves out to sea and let her drown beneath those waves would break the whole point of our betrothal bond.

"Well, I – "

"Ahuja," said Bekka. "I think I'm pregnant."

"What?"

"I missed my last period. And I got sick this morning, even though I didn't eat any extra sparkroot. I think I'm pregnant. I don't want to die. We need to save me. To save the baby."

I looked at the horrible waves, watched them pelt the shore. A few moments in that, and there was no doubt we would capsize and die.

"Okay," I said. "New plan. The three of us go back. We tell them that we couldn't get to Tahiti. They point out that another night has passed. Now four of us have to die. The three of us vote for everybody except Bekka dying. It's 3-2, we win. The rest of us die, and Bekka goes back to the village and the baby lives."

"Hold on," said Daho. "I'm supposed to vote for me to die and Bekka to live? What do I get out of this deal?"

The Tablets of Enku say one man must not kill another. So I didn't.

"You get an extra day!" I snapped. "One extra day of life for saving my betrothed and unborn child. Because we're not going back unless you agree to this. It's either die now, or die tomorrow night. And a lot of things can happen in a day."

"Like what?"

"Like I don't know. We might think of some clever way out. Enku the Lawgiver might return from the dead and change the rules. Whatever. It's a better deal than you'll get if you throw yourself into that water."

Daho glared at me, then weighed his options. "Okay," he snapped. "I'll vote for Bekka. But you had better be thinking *really* hard about those clever ways out."

Day Three

"So," said Calkas the next morning. "I guess all of us are still here." He didn't really sound surprised.

I explained what had happened the night before.

"It's simple," Calkas declared. "The Volcano God is punishing us. He's saying that it's wrong of us to try to escape his judgment by going to Tahiti. That's why he sent the storm. He wants us all to stay here until the bitter end and then, if we have to, we die together."

"No!" I protested. "That's not it at all! The taboo doesn't say we all have to die. It just says we all have to die if we figure out what our eye color is! If some of us kill ourselves, we can prevent that from happening!"

"The Volcano God loathes the needless taking of life," said Calkas. "And he loathes his people traveling to other lands, where the sparkroot never grows and the taboos are violated every day. That's what he's trying to tell us. He's trying to close off our options, so that we stay pure and our souls don't have to burn in his caldera. You know, like Enuli's will." He shot her a poison glance.

"My name is –" she started.

"I don't think that's it at all," I said. "I say the four of us sacrifice ourselves to save Bekka."

"You *would* say that, as her betrothed," said Enuli.

"Well yes," I said. "Yes, I would. Forgive me for not wanting the love of my life to die for a stupid reason. Maybe I should just throw myself in the caldera right now. And she's carrying an unborn child? Did you miss that part?"

"People, people," said Calkas. "Peace! We're all on the same side here."

"No we're not," I said. "So let's vote. Everyone in favor of saving Bekka, say aye."

"And everyone in favor of not sacrificing anyone to the waves, and letting the Volcano God's will be done, say nay."

Calkas added.

"Aye," I said.

"Aye," said Bekka.

"Nay," said Calkas.

"Nay," said Enuli.

"Nay," said Daho.

"What?!" I protested.

"Nay," Daho repeated.

"But you said –" I told him.

"You promised me one extra day," Daho said. "Think about it. Calkas is promising me two."

"No!" I protested. "You can't do this! Seriously, I'll kill you guys if I have to!"

"Then your soul will burn in the caldera forever," said Calkas. "And it still won't help your betrothed or your child."

"You can't do this," I repeated, softly, more of a mutter.

"We can, Ahuja," said Calkas.

I slumped back into my room, defeated.

Day Four

I gave them the traditional morning greeting. "So, I guess we're all still here."

We were. It was our last day. We now had enough information to prove, beyond a shadow of a doubt, that all of us had blue eyes. At midnight, we would all have to commit suicide.

"You know what?" said Enuli. "I've always wanted to say this. ALL OF YOU GUYS HAVE BLUE EYES! DEAL WITH IT!"

We nodded. "You have blue eyes too, Enuli," said Daho. It didn't matter at this point.

"Wait," said Bekka. "No! I've got it! Heterochromia!"

"Hetero-what?" I asked.

"Heterochromia iridum. It's a very rare condition where someone has two eyes of two different colors. If one of us has heterochromia iridum, then we can't prove anything at all! The sailor just said that he saw someone with blue eyes. He didn't say how *many* blue eyes."

"That's stupid, Bekka," Enuli protested. "He said blue eyes, plural. If somebody just had one blue eye, obviously he would have remarked on that first. Something like 'this is the only island I've been to where people's eyes have different colors.'"

"No," said Bekka. "Because maybe all of us have blue eyes, except one person who has heterochromia iridum, and he noticed the other four people, but he didn't look closely enough to notice the heterochromia iridum in the fifth."

"Enuli just said," said Calkas, "that we all have blue eyes."

"But she didn't say how many!"

"But," said Calkas, "if one of us actually had heterochromia iridum, don't you think somebody would have thought to mention it before the fifth day?"

"Doesn't matter!" Bekka insisted. "It's just probabilistic certainty."

"It doesn't work that way," said Calkas. He put an arm on her shoulder. She angrily swatted it off. "Who even decides these things!" she asked. "Why is it wrong to know your own eye color?"

"The eye is the organ that sees," said Calkas. "It's how we know what things look like. If the eye knew what it itself looked like, it would be an infinite cycle, the eye seeing the eye seeing the eye seeing the eye and so on. Like dividing by zero. It's an abomination. That's why the Volcano God, in his infinite wisdom, said that it must not be."

"Well, I know my eyes are blue," said Bekka. "And I don't feel like I'm stuck in an infinite loop, or like I'm an abomination."

"That's because," Calkas said patiently, "the Volcano God, in his infinite mercy, has given us one day to settle our

worldly affairs. But at midnight tonight, we all have to kill ourselves. That's the rule."

Bekka cried in my arms. I glared at Calkas. He shrugged. Daho and Enuli went off together – I guess they figured if it was their last day in the world, they might as well have some fun – and I took Bekka back to our room.

*
*
*

"Listen," I said. "I'm not going to do it."

"What?" she asked. She stopped crying immediately.

"I'm not going to do it. And you don't have to do it either. You should have your baby, and he should have a mother and father. We can wait here. The others will kill themselves. Then we'll go back to the village on our own and say that the rest of them died in the storm."

"But – aren't you worried about the Volcano God burning our souls in his caldera forever?"

"To be honest, I never really paid much attention in Volcano Church. I – I guess we'll see what happens later on, when we die. The important thing is that we can have our child, and he can grow up with us."

"I love you," said Bekka.

"I know," I said.

"I know you know," she said. "But I didn't know that you knew I knew you knew. And now I do."

"I love you too," I said.

"I know," she said.

"I know you know," I said. I kissed her. "I love you and your beautiful blue eyes."

The storm darkened from gray to black as the hidden sun passed below the horizon.

Day Five

"So," I said when the other four had woken up, "I guess all of us are atheists."

"Yeah," said Daho.

"The world is empty and void of light and meaning," said Enuli. "It's the most Goth thing of all."

Calkas sighed. "I was hoping all of you would kill yourselves," he said, "and then I could go home, and my father the high priest would never have to know what happened. I'm sorry for pushing the rest of you. It's just that – if I looked lax, even for a second, he would have suspected, and then I would have been in so much trouble that an eternity in the Volcano God's caldera would look pretty good compared to what would happen when I got back home."

"I think," said Bekka, "that I realized it the first time I ate the sparkroot. Before I'd even finished swallowing it, I was like, wait a second, volcanoes are probably just geologic phenomenon caused by an upwelling of the magma in the

Earth's mantle. And human life probably evolved from primitive replicators. It makes a lot more sense than some spirit creating all life and then retreating to a dormant volcano on some random island in the middle of the nowhere."

"This is great," said Bekka. "Now even if it's a Mersenne prime day I can eat as many green vegetables as I want!"

"You know Mersenne prime days only come like once every couple of centuries, right?" I asked her.

"I know. It's just the principle of the thing."

"We can't tell any of the others," Daho insisted. "They'd throw us into the volcano."

"You think?" I said. "Calkas was saying before that 99% of us had blue eyes, so probably we all had blue eyes. Well, think about it. The five of us are a pretty random sample of the island population, and all five of us are atheist. That means there's probably a lot more. Maybe everybody's atheist."

"Everybody?"

"Well, I thought Calkas was like the most religious of anybody I knew. And here we are."

"I told you, I was just trying to behave so that I didn't get in trouble with my father."

"What if everyone's doing that? Nobody wants to get in trouble by admitting they don't believe, because if anybody else found out, they'd get thrown into the volcano. So we all just put on a mask for everybody else."

"I figured Ahuja was atheist," said Bekka.

"You did?!" I asked her.

"Yeah. It was the little things. When we were hanging out. Sometimes you'd forget some rituals. And then you'd always shoot these guilty glances at me, like you were trying to see if I'd noticed. I thought it was cute."

"Why didn't you tell me?"

"You'd have freaked out. You'd have had to angrily deny it. Unless you knew I was atheist. But I couldn't have told you that, because if I did then you might feel like you had to throw *me* in the volcano to keep up appearances."

"Bekka!" I said. "You know I would never – "

"I kind of suspected Calkas was atheist," said Daho. "He got so worked up about some of those little points of law. It had to be overcompensating."

"Hold on hold on hold on!" said Calkas. "So basically, we were all atheists. We all knew we were all atheists. We just didn't know that we knew that we were all atheists. This is hurting my brain. I think I'm going to need more sparkroot."

A sunbeam peeked through the wall of the lodge.

"Storm's over!" Bekka shouted gleefully. "Time to go back home!" We gathered our things and went outside. The sudden sunlight felt crisp and warm upon my skin.

"So," said Daho, "we don't mention anything about the sailor to anyone else back at the village?"

"Are you kidding?" said Calkas. "I say we stand in the middle of town square, announce everybody's eye colors, and then suggest that maybe they don't believe in the Volcano God as much as they thought. See what happens."

"YOU ALL HAVE BLUE EYES!" Enuli shouted at the jungle around us. "DEAL WITH IT!" We laughed.

"By the way," I told Enuli. "While we're airing out things that everybody knows in order to make them common knowledge, that bone in your hair looks ridiculous."

"He's right," Daho told her.

"It really does," Calkas agreed.

"You watch out," said Enuli. "Now that we don't have to reserve the sparkroot for interpreting taboos, I'm going to invent a death ray. Then you'll be sorry."

"Hey," said Daho, "that sounds pretty cool. And I can invent a giant aerial dreadnaught to mount it on, and together we can take over Europe and maybe the next sailor who gets shipwrecked on our island will be a little less condescending."

"Ha!" said Enuli. "That would be so Goth."

Sun on our backs, we took the winding road into the village.