Slate Star Codex, 2015

# Untitled

EDIT: This is the most controversial post I have ever written in ten years of blogging. I wrote it because I was very angry at a specific incident. I stand by a lot of it, but if somebody links you here saying “HERE’S THE SORT OF GUY THIS SCOTT ALEXANDER PERSON IS, READ THIS SO YOU KNOW WHAT HIS BLOG IS REALLY ABOUT”, please read any other post instead. There’s a whole list of Top Posts on the Top Posts bar above.

Trigger warning: social justice, condemnation of some feminism, tangential reference to eating disorder. Note that although our names are very similar, I am NOT the same person as Scott Aaronson and he did NOT write this article. Not meant as a criticism of feminism, so much as of a certain way of operationalizing feminism.

I.

In my heart, there is a little counter that reads “XXX days without a ten-thousand word rant about feministm.” And I had just broken three digits when they had to go after Scott Aaronson.

For those of you who don’t know, Scott Aaronson is one of the nicest, smartest, and most decent people there are. A few days ago, in response to a discussion of sexual harassment at MIT, Aaronson reluctantly opened up about his experience as a young man:

I check Feministing, and even radfem blogs like “I Blame the Patriarchy.” And yes, I’ve read many studies and task force reports about gender bias, and about the “privilege” and “entitlement” of the nerdy males that’s keeping women away from science. Alas, as much as I try to understand other people’s perspectives, the first reference to my “male privilege”—my privilege!—is approximately where I get off the train, because it’s so alien to my actual lived experience.

But I suspect the thought that being a nerdy male might not make me “privileged”—that it might even have put me into one of society’s least privileged classes—is completely alien to your way of seeing things. To have any hope of bridging the gargantuan chasm between us, I’m going to have to reveal something about my life, and it’s going to be embarrassing.

(sigh) Here’s the thing: I spent my formative years—basically, from the age of 12 until my mid-20s—feeling not “entitled,” not “privileged,” but terrified. I was terrified that one of my female classmates would somehow find out that I sexually desired her, and that the instant she did, I would be scorned, laughed at, called a creep and a weirdo, maybe even expelled from school or sent to prison. You can call that my personal psychological problem if you want, but it was strongly reinforced by everything I picked up from my environment: to take one example, the sexual-assault prevention workshops we had to attend regularly as undergrads, with their endless lists of all the forms of human interaction that “might be” sexual harassment or assault, and their refusal, ever, to specify anything that definitely wouldn’t be sexual harassment or assault. I left each of those workshops with enough fresh paranoia and self-hatred to last me through another year.

My recurring fantasy, through this period, was to have been born a woman, or a gay man, or best of all, completely asexual, so that I could simply devote my life to math, like my hero Paul Erdös did. Anything, really, other than the curse of having been born a heterosexual male, which for me, meant being consumed by desires that one couldn’t act on or even admit without running the risk of becoming an objectifier or a stalker or a harasser or some other creature of the darkness.

Of course, I was smart enough to realize that maybe this was silly, maybe I was overanalyzing things. So I scoured the feminist literature for any statement to the effect that my fears were as silly as I hoped they were. But I didn’t find any. On the contrary: I found reams of text about how even the most ordinary male/female interactions are filled with “microaggressions,” and how even the most “enlightened” males—especially the most “enlightened” males, in fact—are filled with hidden entitlement and privilege and a propensity to sexual violence that could burst forth at any moment.

Because of my fears—my fears of being “outed” as a nerdy heterosexual male, and therefore as a potential creep or sex criminal—I had constant suicidal thoughts. As Bertrand Russell wrote of his own adolescence: “I was put off from suicide only by the desire to learn more mathematics.”

At one point, I actually begged a psychiatrist to prescribe drugs that would chemically castrate me (I had researched which ones), because a life of mathematical asceticism was the only future that I could imagine for myself. The psychiatrist refused to prescribe them, but he also couldn’t suggest any alternative: my case genuinely stumped him. As well it might—for in some sense, there was nothing “wrong” with me. In a different social context—for example, that of my great-grandparents in the shtetl—I would have gotten married at an early age and been completely fine. (And after a decade of being coy about it, I suppose I’ve finally revealed the meaning of this blog’s title.) […]

Now, the whole time I was struggling with this, I was also fighting a second battle: to maintain the liberal, enlightened, feminist ideals that I had held since childhood, against a powerful current pulling me away from them. I reminded myself, every day, that no, there’s no conspiracy to make the world a hell for shy male nerds. There are only individual women and men trying to play the cards they’re dealt, and the confluence of their interests sometimes leads to crappy outcomes. No woman “owes” male nerds anything; no woman deserves blame if she prefers the Neanderthals; everyone’s free choice demands respect.

That I managed to climb out of the pit with my feminist beliefs mostly intact, you might call a triumph of abstract reason over experience. But I hope you now understand why I might feel “only” 97% on board with the program of feminism.

All right. Guy opens up for the first time about how he was so terrified of accidentally hurting women that he became suicidal and tried to get himself castrated. Eventually he got over it and is now 97% on board with feminism, but wants people to understand that when done wrong it can be really scary.

The feminist blogosphere, as always, responded completely proportionally. Amanda Marcotte, want to give us a representative sample?

[Aaronson’s post] is the whole “how can men be oppressed when I don’t get to have sex with all the hot women that I want without having to work for it?” whine, one that, amongst other things, starts on the assumption that women do not suffer things like social anxiety or rejection…It was just a yalp of entitlement combined with an aggressive unwillingness to accept that women are human beings just like men. [He is saying that] “having to explain my suffering to women when they should already be there, mopping my brow and offering me beers and blow jobs, is so tiresome…I was too busy JAQ-ing off, throwing tantrums, and making sure the chip on my shoulder was felt by everyone in the room to be bothered to do something like listen.” Women are failing him by not showing up naked in his bed, unbidden. Because bitches, yo.

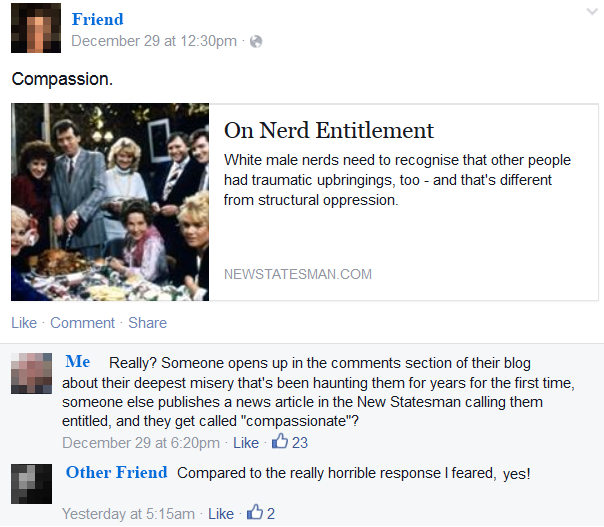
The eternal struggle of the sexist: Objective reality suggests that women are people, but the heart wants to believe they are a robot army put here for sexual service and housework.

This would usually be the point where I state for the record that I believe very strongly that all women are human beings. Problem is, I’ve just conceived a sudden suspicion that one of them is actually a Vogon spy in a skin suit.

Anyway, Marcotte was bad enough, given that she runs one of the most-read feminist blogs on the Internet. But much of the rest of the feminist “discussion” on Tumblr, Twitter, and the like was if anything even worse.

But there was one small ray of hope. A bunch of people sent me an article on the issue by Laurie Penny in New Statesman, called “On Nerd Entitlement: White Male Nerds Need To Recognize That Other People Had Traumatic Upbringings Too And That’s Different From Structural Oppression.” The article was always linked with commentary like “This is so compassionate!” or “Finally a decent human being is addressing this issue with kindness!”

Well, I read the article, and ended up having the following Facebook conversation:



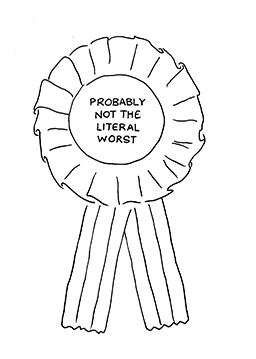
On further reflection, Other Friend has a point. I disliked Penny’s article, but compared to everything else it was a ray of light, a breath of fresh air, an unexpected incursion from a utopia of universal love and understanding. I didn’t feel like it treated Aaronson fairly. But I did feel like it treated him like a human being, which is rare and wonderful.

From the article:

I do not intend for a moment to minimise Aaronson’s suffering. Having been a lonely, anxious, horny young person who hated herself and was bullied I can categorically say that it is an awful place to be. I have seen responses to nerd anti-feminism along the lines of “being bullied at school doesn’t make you oppressed”. Maybe it’s not a vector of oppression in the same way, but it’s not nothing. It burns. It takes a long time to heal.

That this article keeps being praised effusively for admitting that someone else’s suicidal suffering “isn’t nothing”, is a sign. It’s a sign of how low our standards are. But it’s also a sign people are ready for change.

It’s hard for me express simultaneously both how genuinely grateful and impressed I am that the article managed to avoid being awful, and how far I still think it has to go. I can only offer Ms. Penny and the entire staff of the New Statesman the recognition appropriate for their achievement:



I’ve already written some thoughts on this general issue in Radicalizing The Romanceless. But by bringing nerd-dom into the picture, Penny has made that basic picture exponentially more complicated.

Luckily, this is a post about Scott Aaronson, so things that become exponentially more complicated fit the theme perfectly.

II.

Ms. Penny writes:

Feminism is not to blame for making life hell for “shy, nerdy men”. It is a real shame that Aaronson picked up Andrea Dworkin rather than any of the many feminist theorists and writers who manage to combine raw rage with refusal to resort to sexual shame as an instructive tool. Weaponised shame – male, female or other – has no place in any feminism I subscribe to.

I live in a world where feminists throwing weaponized shame at nerds is an obvious and inescapable part of daily life. Whether we’re “mouth-breathers”, “pimpled”, “scrawny”, “blubbery”, “sperglord”, “neckbeard”, “virgins”, “living in our parents’ basements”, “man-children” or whatever the insult du jour is, it’s always, always, ALWAYS a self-identified feminist saying it. Sometimes they say it obliquely, referring to a subgroup like “bronies” or “atheists” or “fedoras” while making sure everyone else in nerddom knows it’s about them too.

There continue to be a constant stream of feminist cartoons going around Tumblr featuring blubberous neckbearded fedora-wearing monsters threatening the virtue of innocent ladies.



Oops, I accidentally included three neo-Nazi caricatures of Jews in there. You did notice, right?

Read any article from the appropriate subfield of feminism, and you may well run into the part with the girl walking into a comic book store only to be accosted by a mouth-breathing troglodyte, followed by a “lesson” on nerd male privilege.

But it’s not just that. Try to look up something on Iron Man, and you get an article on Iron Man-Child and how “the white maleness of geek culture” proves they are “the most useless and deficient individuals in society, precisely because they have such a delusional sense of their own importance and entitlements.” Go to Jezebel and people are talking about how jocks are so much better than nerds because nerds hate women.

It has reached the point where articles published in major journals talk about the the fedora phenomenon in the context of “the growing trend in feminists and other activists online that use shaming as an activist strategy”.

Let’s not mince words. There is a growing trend in Internet feminism that works exactly by conflating the ideas of nerd, misogynist, virgin, person who disagrees with feminist tactics or politics, and unlovable freak.

Ms. Penny may be right that her ideal feminism doesn’t do that. Then again, my ideal masculinity doesn’t involve rape or sexual harassment. Ideals are always pretty awesome. But women still have the right to complain when actual men rape them, and I’m pretty sure nerds deserve the right to complain that actual feminists are, a lot of the time, focused way more on nerd-baiting than actual feminism, and that much the same people who called us “gross” and “fat” and “loser” in high school are calling us “gross” and “misogynist” and “entitled” now, and for much the same reasons.

III.

Penny goes on to deny that this is a gendered issue at all:

Like Aaronson, I was terrified of making my desires known- to anyone. I was not aware of any of my (substantial) privilege for one second – I was in hell, for goodness’ sake, and 14 to boot…Scott, imagine what it’s like to have all the problems you had and then putting up with structural misogyny on top of that. Or how about a triple whammy: you have to go through your entire school years again but this time you’re a lonely nerd who also faces sexism and racism.

This comes across so strongly as “my suffering is worse than your suffering” spiel, so much so that I’m tempted to argue it and review a bunch of experiments like how even the least attractive women on dating sites get far more interest than men. Or how women asking random people for sex on the street get accepted more than two-thirds of the time, but men trying the same get zero percent. Or how the same study shows that the women who get declined get declined politely, while the men are treated with disgust and contempt. Or I could hunt down all of the stories of trans men who start taking testosterone, switch to a more male sex drive, and are suddenly like “OH MY GOD I SUDDENLY REALIZE WHAT MALE HORNINESS IS LIKE I THOUGHT I KNEW SEXUAL FRUSTRATION BEFORE BUT I REALLY REALLY DIDN’T HOW DO YOU PEOPLE LIVE WITH THIS?”

But my commenters have convinced me that taking this further would be joining in the pissing contest I’m condemning, so let’s put it a little differently.

A couple of studies show that average-attractiveness people who ask random opposite-gender strangers on dates are accepted 50% of the time, regardless of their gender.

Grant that everyone involved in this conversation has admitted they consider themselves below average attractiveness (except maybe Marcotte, whose daily tune-ups keep her skin-suit in excellent condition). Fine. Maybe we have a success rate of 10%?

That’s still astounding. It would be pretty easy to mock teenage-me for not asking for dates when ten percent of people would have said yes. Asking ten people something takes what, five minutes? And would have saved how many years of misery?

This is a pretty impressive market failure – in sheer utility cost, probably bigger than any of the market failures actual economists talk about.

Some people say the female version of the problem is men’s fault, and call the behavior involve slut-shaming. I take this very seriously and try not to slut-shame or tolerate those who do.

But the male version of the problem is nerd-shaming or creep-shaming or whatever, and I don’t feel like most women, especially most feminist women, take it nearly as seriously as I try to take their problems. If anything, many actively make it worse. This is exactly those cartoons above and the feminists spreading them. Nerds are told that if they want to date girls, that makes them disgusting toxic blubberous monsters who are a walking offense to womankind.

This is maybe not the most reasonable interpretation of modern sexual mores, but neither is “any women who has sex before marriage is a slut and no one will ever value her.” Feminists are eagle-eyed at spotting the way seemingly innocuous messages in culture can accidentally reinforce the latter, but continue to insist that there’s no possible way that shouting the former from the rooftops could possibly lead to anyone believing or internalizing it.

Talking about “entitled nerds” is the Hot New Internet Feminism thing these days. Here’s The Entitlement And Misogyny Of Nerd Culture. Here’s Sex, Nerds, Entitlement, and Rape. Here’s Is Nerd Culture Filled With Entitled Crybabies? There’s On Male Entitlement: Geeks, Creeps, and Sex.

And now, apparently, the New Statesman, realizing that it’s almost 2015 and it has yet to claim a share of the exciting nerd entitlement action, has On Nerd Entitlement by Laurie Penny

And this is more than a little weird, because the actual nerds I know in real life tend to be more like Scott Aaronson, who is spending less time feeling entitled to sex, and more time asking his doctor if there’s any way to get him castrated because his sexual desire might possibly offend a woman. Or more like me, who got asked out by a very pretty girl in middle school and ran away terrified because he knew nobody could actually like him and it was obviously some kind of nasty trick.

So given that real-life nerds are like this, and given that they’re sitting around being terrified that they’re disgusting toxic monsters whose wish to have sex is an offense against womankind, what do you think happens when they hear from every news source in the world that they are entitled?

What happens is they think “Oh God! There was that one time when I looked at a woman and almost thought about asking her out! That means I must be feeling entitled to sex! I had temporarily forgotten that as a toxic monster I must never show any sexuality to anybody! Oh God oh God I’m even worse than I thought!”

Again, this is not the most rational thing in the world. But I maintain it’s no less rational than, say, women who won’t leave their abusive husband because he’s convinced them they don’t deserve anything better than what they get. Gender is weird. Self-loathing is easy to inculcate and encourage, even unintentionally. Heck, we’ve already identified this market failure of people preferring to castrate themselves rather than ask ten people on a date, something weird has got to explain it.

When feminists say that the market failure for young women is caused by slut-shaming, I stop slut-shaming, and so do most other decent people.

When men say that the market failure for young men is caused by nerd-shaming, feminists write dozens of very popular articles called things like “On Nerd Entitlement”.

The reason that my better nature thinks that it’s irrelevant whether or not Penny’s experience growing up was better or worse than Aaronson’s: when someone tells you that something you are doing is making their life miserable, you don’t lecture them about how your life is worse, even if it’s true. You STOP DOING IT.

IV.

This also serves to illuminate what I think is the last and most important difference between Penny’s experience and Aaronson’s experience.

When Penny bares her suffering to the world for all to hear about, she gets sympathy, she gets praised as compassionate, she gets published in important magazines whose readers feel sorry for her and acknowledge that her experience sucks.

When Aaronson talks about his suffering on his own blog, he gets Amanda Marcotte. He gets half the internet telling him he is now the worst person in the world.

This was my experience as well. When I complained that I felt miserable and alone, it was like throwing blood in the water. A feeding frenzy of feminists showed up to tell me I was a terrible person and deserved to die, sometimes in terms that made Marcotte look like grandmotherly kindness. This is part of the experience I write about in this post, and it’s such a universal part of the shy awkward male experience that we are constantly flabbergasted that women refuse to accept it exists.

When feminists write about this issue, they nearly always assume that the men involved are bitter about all the women who won’t sleep with them. In my experience and the experience of everyone I’ve ever talked to, we’re bitter about all the women who told us we were disgusting rapists when we opened up about our near-suicidal depression.

And when that happens, again and again and again, of course we learn to shut up about it. I bottled my feelings inside and never let them out and spent years feeling like I was a monster for even having them.

As a mental health professional, I can assure you this is the best coping strategy.

V.

Laurie Penny has an easy answer to any claims that any of this is feminists’ fault:

Feminism, however, is not to blame for making life hell for “shy, nerdy men”. Patriarchy is to blame for that.

I say: why can’t it be both?

Patriarchy is yet another motte and bailey trick.

The motte is that patriarchy is the existence of different gender roles in our society and the ways in which they are treated differently.

The bailey is that patriarchy is men having power over women.

If you allow people to switch between these and their connotations willy-nilly, then you enable all sorts of mischief.

Whenever men complain about anything, you say “Oh, things are bad for men? Well, that sounds like a gender role. Patriarchy’s fault!”

And then the next day you say “Well, since we already agreed yesterday your problem is patriarchy, the solution is take away power from men and give it to women. It’s right there in the word, patri-archy. So what we need is more feminism.”

Even if in this particular case the feminism is making the problem worse.

So, for example, we are told that the patriarchy causes male rape. We are told that if we want to fight male rape, the best way to do so is to work hard to promote feminist principles. But once feminism has been promoted, the particular feminists benefitting from that extra social capital may well be the ones to successfully lobbying national governments to keep male rape legal on the ground that if raping men was illegal, they might make false accusations which could hurt women.

If patriarchy is “any problem with gender roles”, it’s entirely possible, even predictable, that feminists can be the ones propping it up in any given situation.

I mean, we live in a world where the Chinese Communist Party is the group that enforces Chinese capitalism and oppresses any workers who complain about it. We live in a world where the guy who spoke out against ritualized purity-obsessed organized religion ended up as the founder of the largest ritualized purity-obsessed organized religion of all time. We live in a world where the police force, which is there to prevent theft and violence, is confiscating property and shooting people right and left. It seems neither uncommon nor unexpected that if you charge a group with eliminating an evil that’s really hard to eliminate, they usually end up mildly tweaking the evil into a form that benefits them, then devoting most of their energy to punishing people who complain.

Pick any attempt to shame people into conforming with gender roles, and you’ll find self-identified feminists leading the way. Transgender people? Feminists led the effort to stigmatize them and often still do. Discrimination against sex workers? Led by feminists. Against kinky people? Feminists again. People who have too much sex, or the wrong kind of sex? Feminists are among the jeering crowd, telling them they’re self-objectifying or reinforcing the patriarchy or whatever else they want to say. Male victims of domestic violence? It’s feminists fighting against acknowledging and helping them.

Yes, many feminists have been on both sides of these issues, and there have been good feminists tirelessly working against the bad feminists. Indeed, right now there are feminists who are telling the other feminists to lay off the nerd-shaming. My girlfriend is one of them. But that’s kind of my point. There are feminists on both sides of a lot of issues, including the important ones.

(“But nowadays in 2015 most feminists are on the right side of every gender issue, right?” Insofar as your definition of ‘the right side of a gender issue’ is heavily influenced by ‘the side most feminists are on’, I’m going to have a really hard time answering that question in a non-tautologous way. Come back in 2065 and we can have a really interesting discussion about whether the feminists of 2015 screwed up as massively as the feminists of 1970 and 1990 did.)

So feminists can be either against or in favor of “patriarchy” broadly defined. Whether or not a form of cruelty is decreed to be patriarchy doesn’t tell us how many feminists are among the people twisting the knife.

The preferred method of figuring this out is asking the people involved.

I’ve been saying for years that getting exposed to feminist shaming was part of what made my adolescence miserable. Every time I say this, I get a stream of grateful emails thanking me for saying something so true to their experience.

Scott Aaronson has now said that getting exposed to feminist shaming was part of what made his adolescence miserable. According to his most recent blog post, he’s also getting the stream of grateful emails:

Throughout the past two weeks, I’ve been getting regular emails from shy nerds who thanked me profusely for sharing as I did, for giving them hope for their own lives, and for articulating a life-crushing problem that anyone who’s spent a day among STEM nerds knows perfectly well, but that no one acknowledges in polite company. I owe the writers of those emails more than they owe me, since they’re the ones who convinced me that on balance, I did the right thing.

I hang out a lot with shy awkward nerdy men of all ages, and I very often hear from them that feminist shaming is part of what’s making their adolescence (and often current life) miserable.

And it’s not just men. Here’s what a lesbian friend of mine had to say about Penny’s article:

There are a hell of a lot of people attracted to women who seem to have internalized the message that their attraction makes them sick and wrong and evil and creepy, that basically any interaction they have with a woman is coercive or harmful on their part, and that initiating a romantic interaction makes them a sexual predator.

I know this because I’m one of them.

I’m a woman. I’m gay. By the time I realized that second thing, I’d internalized that all attraction to women was objectifying and therefore evil. I spent years of my life convinced that it was coercive to make it clear to girls that I wanted to date them, lest they feel pressured. So I could only ask them out with a clear conscience if I was in fact totally indifferent to their answer. I still decide I’m abusive pretty frequently, on the basis of things like ‘i want to kiss her, which is what an abuser would want’ and ‘i want to be special to her, which is what an abuser would want’.

I internalized these messages from exposure to feminist memes, norms, and communities. It was feminist messages, not homophobic ones, that made it hardest for me to come to terms with my sexuality. It wasn’t intentional. But it happened. And it has happened by now to enough people that ‘well obviously you’re misinterpreting it’ is starting to wear thin as an excuse. Lots and lots of people are misinterpreting the way I did. By and large, we’re vulnerable people. Very often we’re mentally ill or disabled people.

Even if it’s broadly good for feminism to emphasize narratives about objectification and entitlement, this seems like a negative consequence of the way contemporary feminist activism does that. Activism shouldn’t make vulnerable people suicidally guilty. If there was a way to do activism that didn’t have this consequence, it’d be better than the current setup.

The infuriating thing is that I think there might be. We could write articles acknowledging that certain conversations can exacerbate crippling guilt and self-loathing, particularly for people with anxiety, depression, or other mental illnesses that make them fixate on their own perceived worthlessness. We could really, truly, not-just-lip-service integrate concern for those people into our activism. We could acknowledge how common this experience is and have resources to help people. We could stop misidentifying anguish as entitlement, and stop acting like anguish that does have entitlement at its root is deserved or desirable or hilarious.

We could really just start by extending to men who share this experience with me the sympathy that I’m extended when I talk about it.

The responses on Tumblr from men and women all over the sexuality spectrum who have had any personal experience with this all say it’s how they feel as well.

I usually avoid the term “privilege” because it tends to start World War III when used. So let’s avoid the term and simply keep in mind the concept that people have private information about their own experience that it’s difficult for other people to get second-hand.

Ms. Penny, as an (I think?) heterosexual woman, has no idea what having to deal with our culture’s giant minefield around romance toward women is like.

Scott Aaronson is a straight guy, and he’s saying feminist shaming tactics have made it worse. I’m an asexual heteroromantic guy, and I’m telling her feminist shaming tactics have made it worse. Unitofcaring is a lesbian woman, and she’s saying feminist shaming tactics have made it worse. HughRistik, who is some sort of weird metrosexual something (I mock him because I love him), is telling her feminist shaming tactics have made it worse. A giant cry has arisen from shy awkward men, lesbians, bisexuals, whatever of the world is saying “NO, SERIOUSLY, FEMINIST SHAMING TACTICS ARE MAKING THIS WORSE”

When Ms. Penny protests that feminism can’t possibly be involved and all these other people’s s personal experience is wrong, this is coming from a place of startling arrogance. If patriarchy means everything in the world, then yes, it is the fault of patriarchy. But it’s the kind of patriarchy that feminism as a movement is working day in and day out to reinforce.

VI.

The subtitle of the article is “White male nerds need to recognise that other people had traumatic upbringings, too – and that’s different from structural oppression.”

This doesn’t really describe the argument very well. The closest it really comes is to say that:

Aaronson makes a sudden leap, and it’s a leap that comes right from the gut, from an honest place of trauma and post-rationalisation, from that teenage misery to a universal story of why nerdy men are in fact among the least privileged men out there, and why holding those men to account for the lack of representation of women in STEM areas – in the most important fields both of human development and social mobility right now, the places where power is being created and cemented right now – is somehow unfair […]

This is why Silicon Valley is fucked up. Because it’s built and run by some of the most privileged people in the world who are convinced that they are among the least.

I really fucking hope that it got better, or at least is getting better, At the same time, I want you to understand that that very real suffering does not cancel out male privilege, or make it somehow alright. Privilege doesn’t mean you don’t suffer, which, I know, totally blows.

The impression I’m getting is that yes, nerds think they have problems, but actually they’re really privileged. So their problems aren’t structural oppression in the same sense that women’s problems are. So. Quick hypothetical.

I’ve postulated before that “privilege” is a classic motte-and-bailey term. The motte, the uncontroversial and attractive definition, is “some people have built-in advantages over other people, and it might be hard for them to realize these advantages even exist”. Under this definition, it’s easy to agree that, let’s say, Aaronson has the privilege of not having to deal with slut-shaming, and Penny has the privilege of not having to deal with the kind of creep-shaming that focuses on male nerds.

The bailey, the sneaky definition used to push a political point once people have agreed to the motte, is that privilege is a one-dimensional axis such that for any two people, one has privilege over the other, and that first person has it better in every single way, and that second person has it worse in every single way.

This is of course the thing everyone swears they don’t mean when they use the word privilege, which is of course how the motte-and-bailey fallacy works. But as soon as they are not being explicitly challenged about the definition, this is the way they revert back to using the word.

Go back to the original Amanda Marcotte article. Check the title. “MIT Professor Explains The Real Oppression Is Having To Talk To Women”.

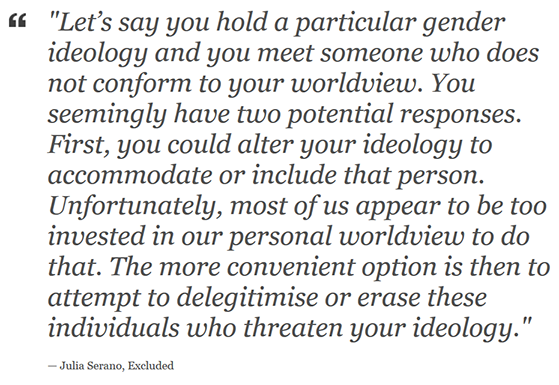
That phrasing, “the real oppression is…”, carries a pretty loaded assumption. I’d say “hides a pretty loaded assumption”, but it doesn’t seem to be doing much work to hide it.

If you look through Marcotte’s work, you find this same phrasing quite often. “Some antifeminist guy is ranting at me about how men are the ones who are really oppressed because of the draft” (source). And she’s not the only one. If you Google the term “are the ones who are really oppressed”, you can find an nice collection of people using this exact phraseology, including a few examples from a charming site called “Nerds Fucking Suck”.

But Aaronson is admitting about a hundred times that he recognizes the importance of the ways women are oppressed. He’s not saying his suffering is worse than women’s in every way, just that it’s really bad and maybe this is not the place where “male privilege” should be invoked. The “is really oppressed” isn’t taken from him, it’s assumed by Marcotte. Her obvious worldview is – since privilege and oppression are a completely one dimensional axis, for Aaronson to claim that there is anything whatsoever that has ever been bad for men must be interpreted as a claim that they are the ones who are really oppressed and therefore women are not the ones who are really oppressed and therefore nothing whatsoever has ever been bad for women. By Insane Moon Logic, it sort of makes sense.

As a result, Marcotte is incapable of acknowledging that Aaronson feels pain or has feelings more complicated than “all women exist solely to be my slaves”. She has to be a jerk to him, otherwise it would be a tacit admission that he has problems, which means only he has problems, which means no woman has ever had problems, which means all women are oppressors. Or whatever.

Marcotte is angry that Aaronson doesn’t cite any feminist writer besides Andrea Dworkin, so let’s go with Julia Serano here:



What if you’re trying to hold the same weird one-dimensional system in a way consistent with basic human decency? That is, you don’t want to do the Vogon thing and say Scott Aaronson’s misery is totally hilarious, but you also don’t want to acknowledge that it counts – because if it counted you’d have to admit that men have it bad in some ways, which means that the One Group That Can Ever Have Things Bad spot is taken by men, which means women don’t have it bad?

As best I can tell, the way with the fewest epicycles is to say “Yes, your pain technically exists, but it’s not structural oppression“, where structural oppression is the type of pain that fits neatly onto the one-dimensional line.

Laurie Penny is an extremely decent person, but like a shaman warding off misfortune with a ritual, she must dub Aaronson’s pain “not structural oppression” or else risk her own pain not counting, being somehow diminished.

I mean, I don’t think she thinks that’s what she’s doing. But I’m not sure why else it’s necessary to get so competitive about it.

Absent the one-dimensional view, it would be perfectly reasonable to say something like “You feel pain? I have felt pain before too. I’m sorry about your pain. It would be incredibly crass to try to quantify exactly how your pain compares to my pain and lord it over you if mine was worse. Instead I will try to help you with your pain, just as I hope that you will help me with mine.”

Given the one-dimensional view, any admission that other people suffer is a threat to the legitimacy of one’s own suffering. Horrible people will deny and actively mock the pain of others, but even decent people will only be able to accept the pain if they also mention in an aside that it doesn’t count as the correct sort of pain to matter in the moral calculus and certainly isn’t even in the same ballpark as their own.

But the one-dimensional view sucks. It is the culmination and perfection of the phenomenon I described in my post on social justice terminology, the abandonment of discourse about the world in favor of endless debate about who qualifies for certain highly loaded terms like “structural oppression”. And those terms end up as a sort of Orwellian Newspeak that makes it possible to dismiss entire categories of experience and decree by fiat who does and doesn’t matter.

Imagine a boot stamping on a human face forever, saying "I KNOW YOU FEEL UPSET RE STAMPING, BUT THAT'S DIFFERENT FROM STRUCTURAL OPPRESSION"

— Scott Alexander (@slatestarcodex) December 31, 2014

The boot acknowledged my pain! So compassionate!

§

The suspect famously says “I didn’t kill him, officer! Also, he had it coming!”

In that spirit, I would like to propose that we shouldn’t make this debate about structural oppression, but even if we do this kind of minimization of male nerd suffering doesn’t stand.

I know there are a couple different definitions of what exactly structural oppression is, but however you define it, I feel like people who are at much higher risk of being bullied throughout school, are portrayed by the media as disgusting and ridiculous, have a much higher risk of mental disorders, and are constantly told by mainstream society that they’re ugly and defective kind of counts.

If nerdiness is defined as intelligence plus poor social skills, then it is at least as heritable as other things people are willing to count as structural oppression like homosexuality (heritability of social skills, heritability of IQ, heritability of homosexuality) If all nerds were born with blue dots on their heads, and the blue-dotters were bullied in school, cast negatively in the media, assumed to be as ravenous beasts hungry for innocent women, and denounced as “entitled” any time they overcame all this to become successful – would anybody deny that blue-dotters suffered from structural oppression? Wouldn’t the people who talked about how clearly blue-dotters are entitled dudebros in the tech industry be thought of the same way as someone who said Jews were greedy parasites in the banking industry?

Actually, let’s take this Jew thing and run with it. I am not the first person to notice that there are a lot of Jews in Silicon Valley. By maternal descent, at least Mark Zuckerberg, Sergey Brin, Larry Page, Michael Dell, Steve Ballmer, Larry Ellison, and Sheryl Sandberg.

Imagine how an anti-Semite might think about this. “Jews say they’re oppressed. But actually they’re all rich. Oppression disproved!”

In fact, he might he add exactly the same comment we see in the Statesman article: “This is why Silicon Valley is fucked up. Because it’s built and run by some of the most privileged people in the world who are convinced that they are among the least.”

But once again this only works when you have the dumb one-dimensional model of privilege. Some Jews are rich, therefore all Jews are rich, therefore all Jews are privileged, therefore no Jew could be oppressed in any way, therefore Jews are the oppressors.

And much the same is true of nerds. In fact, have you noticed actual nerds and actual Jews tend to be the same people? I’m Jewish. Scott Aaronson is Jewish. Laurie Penny, who declares her nerd-girl credentials, is Jewish. We’re discussing a blog called, of all things, Shtetl-Optimized. A minority that makes up 1% of the Anglosphere also makes up three of the three nerds in this conversation. Probability of this happening by chance is (\*calculates\*) exactly one in a million. Aside from Zuckerberg, Page, and all the other famous people, about 40% of top programmers are Jewish.

Judaism and nerdity are not exactly the same, but they sure live pretty close together.

And this is why it’s distressing to see the same things people have always said about Jews get applied to nerds. They’re this weird separate group with their own culture who don’t join in the reindeer games of normal society. They dress weird and talk weird. They’re conventionally unattractive and have too much facial hair. But worst of all, they have the chutzpah to do all that and also be successful. Having been excluded from all of the popular jobs, they end up in the unpopular but lucrative jobs, for which they get called greedy parasites in the Jews’ case, and “the most useless and deficient individuals in society” in the case of the feminist article on nerds I referenced earlier.

Propaganda against the Jews is described as follows:

Since Jews were ugly, they depended on reprehensible methods of sexual conquest. Non-violent means such as money were common, but also violence. Streicher specialized in stories and images alleging Jewish sexual violence. In a typical example, a girl cowers under the huge claw-like hand of a Jew, his evil silhouette in the background. The caption at the bottom of the page: “German girls! Keep away from Jews!” These images were particularly striking and consistent with the larger theme. Although Jews were too cowardly to engage in manly combat and too disgusting to be physically attractive to German women, they were eager to overpower and rape German women, thereby corrupting the Aryan racial stock.

I already know the same machine that turned Aaronson’s “I am 97% on board with feminism” into “I think all women should be my slaves” is focusing its baleful gaze on me. So let me specify what I am obviously not saying. I am not saying nerds have it “just as bad as Jews in WWII Germany” or any nonsense like that. I am not saying that prejudice against nerds is literally motivated by occult anti-Semitism, or accusing anyone of being anti-Semitic.

I am saying that whatever structural oppression means, it should be about structure. And the structure society uses to marginalize and belittle nerds is very similar to a multi-purpose structure society has used to belittle weird groups in the past with catastrophic results.

There is a well-known, dangerous form of oppression that works just fine when the group involved have the same skin color as the rest of society, the same sex as the rest of society, and in many cases are totally indistinguishable from the rest of society except to themselves. It works by taking a group of unattractive, socially excluded people, mocking them, accusing them of being out to violate women, then denying that there could possibly be any problem with these attacks because they include rich people who dominate a specific industry.

[EDIT 1/3: Penny’s same article was reprinted at New Republic, which I guess also realized it gotten a piece of the Hot New Nerd Entitlement Trend yet. Their title was “Nerd Entitlement Lets Men Ignore Racism And Sexism”, which is kind of weird, since Penny’s article doesn’t do anything close to argue for that. Also since surveys show nerd men are more likely to be concerned about racism and sexism than other men – see for example this survey where nerds are far more feminist than average, so much so that nerd men are more feminist than non-nerd women, and since Penny’s article makes nothing even resembling an argument for this position. Once again, this only makes sense if you assume a one-dimensional zero-sum model of privilege, where the fact that miserable male nerds are concentrating on their own desire for the release of death, instead of what women think they should be concentrating on, means they must be universally denying women can have problems.]

[EDIT 1/3, Part 2: New Republic has changed their title. You can still see it in the URL, though]

VII.

It gets worse.

What can I say? This is a strange and difficult age, one of fast-paced change and misunderstandings. Nerd culture is changing, technology is changing, and our frameworks for gender and power are changing – for the better. And the backlash to that change is painful as good, smart people try to rationalise their own failure to be better, to be cleverer, to see the other side for the human beings they are. Finding out that you’re not the Rebel Alliance, you’re actually part of the Empire and have been all along, is painful.

She links this last sentence to an article called Why Nerd Culture Must Die, which, I don’t know, kind of makes me a little more skeptical of all of her protestations that she’s exactly as much of a nerd as anyone else and likes nerds and is really working for nerds’ best interests. The article repeats that nerds think they’re “the Rebel Alliance” but actually are “the Empire”. Ha ha! Burn!

You may be wondering whether you missed the part of Star Wars where Darth Vader is so terrified of hurting or offending other people that he stops interacting with anybody and becomes suicidally depressed for years. Finally, Vader mentions this fact in the comments section of a blog about obscure Sith rituals. The brave Rebel Alliance springs into action and gets all of the Coruscant newspapers to publish articles on how Vader is entitled and needs to check his privilege.

I don’t know. Maybe this was one of those things that got taken out in the Special Edition?

(Han shot first!)

But there’s actually something even creepier going on here which may or may not be intentional.

The Transsexual Empire is a very famous book from the late 1970s subtitled “The Making Of The She-Male” in which feminist activist Janice Raymond argues that transsexuals, despite claiming to be persecuted, form an evil empire dedicated to the reinforcing of patriarchy. It contains delightful passages such as “All transsexuals rape women’s bodies by reducing the real female form to an artifact, appropriating this body for themselves”. The Transgender Studies Reader says that the book “did not invent anti-transsexual prejudice, but it did more to justify and perpetuate it than perhaps any other book ever written.” The response, written by a prominent transgender activist, was titled The Empire Strikes Back – an obvious reference to the Star Wars film published around that time.

So the question is – how come various feminists keep independently choosing the Empire as a metaphor for their enemies?

Once again the one-dimensional model of privilege rears its ugly head.

Transsexuals claimed to be suffering. This was a problem, because some of them were transwomen who had started with the male gender role. They had privilege! And they claimed to be suffering! The one-dimensional model of privilege lifts its eyebrows quizzically and emits a “…wha?”

The solution is to deny their suffering. Not only deny their suffering, but accuse them of being out to “rape women’s bodies”. Not only deny their suffering and accuse them of being rapists, but to insist that they are privileged – no, super-privileged – no, the most privileged – no, a giant all-powerful all-encompassing mass of privilege that controls everything in the world,.

So they became an Empire. How better to drive home the fact that they’re definitely powerful and oppressive and definitely definitely not suffering? Because if they were suffering, it would mean we weren’t.

There’s another word the radical feminists like to use about transsexuals. “It’s aggrieved entitlement,” Lierre Keith tells the New Yorker. “They are so angry that we will not see them as women.” The article continues to explain how “When trans women demand to be accepted as women they are simply exercising another form of male entitlement.”

And sigh, now here come the male nerds and say they’re suffering too, not as much as the transpeople but still a nonzero amount of pain! Is there no end to people who are not us, suffering in inconvenient ways? They say that when they feel haunted by scrupulosity, that shaming them all the time actually makes the problem worse! We need to establish that they’re privileged right away! So how better to rub in the concept of very privileged people than to draw in the old Empire analogy, right? Maybe try the “entitlement” claim again as well? Second time’s the charm!

But let’s be clear. There is a Star Wars metaphor to be made here.

Chancellor Palpatine is, by universal agreement, a great guy. According to Count Dooku, he “speaks honestly and champions the underprivileged” (direct quote from source). But sometimes people get in the way of his mission of helping the underprivileged, and then he has to, you know, tell it like it is.

Like the Senate. When the Senate is not sure they want to hand over power to the Chancellor, he declares that they are corrupt and oppose democracy.

Or the Jedi. When the Jedi resist his rule, he declares that they are obsessed with “gain[ing] power” and “if they are not all destroyed, it will be civil war without end.”

Whenever he wants to steamroll over someone, Palpatine’s modus operandi is to convince everyone that they are scary oppressors. This isn’t just my personal interpretation. Indeed, in Order 66, Palpatine says straight out:

“Beings believe what you tell them. They never check, they never ask, they never think…Tell them you can save them, and they will never ask—from what, from whom? Just say tyranny, oppression, vague bogeymen.”

If we’ve learned anything from the Star Wars prequels, it’s that Anakin Skywalker is unbearably annoying. But if we’ve learned two things from the Star Wars prequels, it’s that the easiest way to marginalize the legitimate concerns of anyone who stands in your way is to declare them oppressors loud enough to scare everyone who listens.

And if the people in the Star Wars universe had seen the Star Wars movies, I have no doubt whatsoever that Chancellor Palpatine would have discredited his opponents by saying they were the Empire.

(seriously, you wanted to throw the gauntlet down to lonely male nerds, and the turf you chose was Star Wars metaphors? HOW COULD THAT POSSIBLY SEEM LIKE A GOOD IDEA?)

VIII.

Unlike Aaronson, I was also female, so when I tried to pull myself out of that hell into a life of the mind, I found sexism standing in my way. I am still punished every day by men who believe that I do not deserve my work as a writer and scholar. Some escape it’s turned out to be.

Science is a way that shy, nerdy men pull themselves out of the horror of their teenage years. That is true. That is so. But shy, nerdy women have to try to pull themselves out of that same horror into a world that hates, fears and resents them because they are women

Scott, imagine what it’s like to have all the problems you had and then putting up with structural misogyny on top of that.

Ms. Penny believes that, as a woman, she’s been unfairly excluded from the life of the mind and, indeed, from every pursuit she might enjoy or use as an escape.

There is something to be discussed here, but I am having trouble isolating Ms. Penny’s exact claim.

“Unfairly excluded from the life of the mind” might suggest she didn’t have the same opportunities as men to participate in higher education, but in fact women are now 33% more likely than men to earn college degrees and women get higher grades in college than men do. They also get well above half of all master’s degrees, and just a slice over half of all Ph.Ds (and rising). Their likelihood of becoming professors is nicely predicted by the percent of degrees they earn at a couple decade interval. The articles about the world of higher education now all have titles like Missing Men or Why Are Men Falling Behind.

Industry isn’t a good example here either. Women in her demographic group – twenty-something and childless – out-earn their male counterparts by almost ten cents on the dollar.

And she’s probably not talking about science, since women earn 55% of science degrees nowadays. They are somewhat overrepresented even in some “hard” sciences like biology, but overwhelmingly so in the social sciences. Over seventy five percent of psychology majors are female – a disproportionate which blows out of the water the comparatively miniscule 60-40 disproportion favoring men in mathematics.

(Hi! Male psychology major here, can confirm!)

When Penny says she as a woman is being pushed down and excluded from every opportunity in academic life, she means that women in a very small subset of subjects centered around computer science and engineering face a gender imbalance about as bad as men do in another collection of subjects such as psychology and education.

Penny attacks nerds for believing that “holding men to account for the lack of representation of women in STEM areas…is somehow unfair.” Fine. I hold her to account for the even higher imbalance in favor of women in psychology and education. Once she accepts responsibility for that, I’ll accept responsibility for hers. That sounds extremely fair.

(“But that’s because of patriarchy!” READ SECTION V.)

I propose an alternate explanation to both dilemmas.

By late high school, the gap between men and women in math and programming is already as large as it will ever be. Yes, it’s true that only 20 – 23% of tech workers are women. But less than twenty percent of high school students who choose to the AP Computer Science test are women.

Nothing that happens between twelfth grade and death decreases the percent of women interested in computer science one whit.

I have no hard numbers on anything before high school, but from anecdotal evidence I know very very many young men who were programming BASIC on their dad’s old computer in elementary school, and only a tiny handful of young women who were doing the same.

I don’t want to get into a drawn out inborn-ability versus acculutration fight here. I want to say that I want to say that whether we attribute this to inborn ability or to acculturation, the entire gender gap has been determined in high school if not before. If anything, women actually gain a few percentage points as they enter Silicon Valley.

What the heck do high schoolers know about whether Silicon Valley culture is sexist or not? Even if you admit that all the online articles talking about this are being read by fourteen year olds in between Harry Potter and Twilight, these articles are a very new phenomenon and my stats are older than they are. Are you saying the is because of a high level of penetration of rumors about “toxic brogrammers” into the world of the average 11th grader?

The entire case for Silicon Valley misogyny driving women out of tech is a giant post hoc ergo propter hoc.

What’s worse, I have never heard any feminist give this case in anything like a principled way. The explanation is usually just something like of course men would use their privilege to guard a well-paying and socially prestigious field like programming from women, men have always guarded their privileges, they’ve never given anything up to women without a fight, etc.

My own field is medicine. More than half of medical students are female. In two years, more than half of doctors in the UK will be female, and the US is close behind.

Medicine is better-paying and more prestigious than programming. It’s also terrible. Medicine is full of extremely abrasive personalities. Medicine has long work hours. Medicine will laugh at you hysterically if you say you want to balance work and family life.

But women can’t get into medicine fast enough. Every so often medical journals and the popular news run scare stories about how there are so many women in medicine now that if they take off time to raise kids at their accustomed rates we’re suddenly going to find ourselves pretty much doctorless.

So any explanation of the low number of women in Silicon Valley has to equally well explain their comparatively high numbers in medicine.

Given all this, it’s really easy for me to see why it’s tempting to blame nerds. Look at these low-status people. It’s their fault. We already dislike them, now we have an even better reason to dislike them that nicely wraps up an otherwise embarassing mystery. They’re clearly repelling women with their rapey creepishness. It doesn’t hurt that occasional high profile stories of sexual harassment come out of Silicon Valley aren’t hard to find and bring viral.

(no one ever asks whether there are an equally high number of stories of sexual harassment in medicine – or law, or any other field – that no one had a reason to publicize. When I was in medical school, there was an extremely creepy incident of sexual harassment/borderline attempted rape involving a female medical student and male doctor at an outlying hospital where I worked. Nobody put it on the front page of Gawker, because the doctor involved wasn’t a nerd and no one feels any particular need to tar all doctors as sexist.)

But again, you really can’t blame this one on Silicon Valley nerds, unless they are breaking into high schools and harassing the women there. And possibly breaking into grade schools, demanding the young boys start tinkering with BASIC. Time for a better theory.

A look at percent female physicians by subspecialty is instructive. The specialty with the most women is pediatrics, followed by child psychiatry, followed by obstetrics, followed by – you get the picture. The specialties with the least women are the various surgeries – the ones where your patient is immobilized, anaesthetized, opened up, and turned into a not-quite-color-coded collection of tubes and wires to poke and prod at – the ones that bear more than a passing resemblance to engineering.

(surgeons are the jockiest jocks ever to jock, so you can’t blame us for this one)

It seems really obvious to me that women – in high schools and everywhere else – have a statistical predilection to like working with people (especially children) and to dislike working with abstract technical poking and prodding. This is a bias clearly inculcated well before SATs and AP exams, one that affects medics and programmers alike.

It’s a bias that probably has both cultural and biological origins. The cultural origins are far too varied to enumerate. Many people very justly bring up the issue of how our society genders toys, with parents getting very angry when girls play with stereotypically male toys and vice versa. The classic example is of course the talking Barbie who would famously say “Math is hard! Let’s go shopping!”

On the other hand, I also think people who neglect biological causes are doing the issue a disservice. Did you know that young monkeys express pretty much exactly the same gendered toy preferences as human children? Rhesus monkeys, vervet monkeys, pretty much whatever species of monkeys you try it on, the male monkeys enjoy wheeled toys more and the female monkeys plush toys more. The word reviewers use to describe the magnitude of the result is “overwhelming”. When intersex children are raised as other than their biological gender, their toy preference and behavior are consistently that associated with their biological gender and not the gender they are being raised as, even when they themselves are unaware their biological gender is different. This occurs even when parents reinforce them more for playing with their gender-being-raised-as toys. You can even successfully correlate the degree of this with the precise amount of androgen they get in the womb, and if you experimentally manipulate the amount of hormones monkeys receive in the womb, their gendered play will change accordingly. 2D:4D ratio, a level of how much testosterone is released during a crucial developmental period, accurately predicts scores both on a UK test of mathematical ability at age seven and the SATs in high school.

The end result of all this is probably our old friend gene-culture interaction, where certain small innate differences become ossified into social roles that then magnify the differences immensely. As a result, high school girls are only a fifth as likely to be interested in computer science as high school boys, and sure enough women are only a fifth as well represented in Silicon Valley as men.

All of this information is accessible for free to anyone who spends ten minutes doing a basic Google search. But instead we have to keep hearing how nerds are gross and disgusting and entitled and should feel constant shame for how they bully and harass the poor female programmers out of every industry they participate in. Penny blames nerds for not “holding men to account for the lack of representation of women in STEM areas” but SERIOUSLY WE DIDN’T DO IT.

(except insofar as we helped acculturate kids. But that’s hardly a uniquely male pasttime.)

(before you bring up that one paper that showed research leaders advantaged male over female researchers, keep in mind that first of all it explains only a small portion of the discrepancy, and second of all the female research leaders showed the bias even worse than the male ones. Yet Penny frames her question as “holding men to account”. This is that motte-and-bailey thing with patriarchy again.)

Do you realize how unpleasant it is to be constantly blamed all the time for something we didn’t do, and have that be used to justify every form of insult and discrimination and accusation against us? The oldest pattern in human history is “Here’s a problem. And here’s a bunch of people who are different than us. Let’s blame it on them!”

There’s enough information out there to prove that creepy nerds are not the problem with female representation in STEM. Then again, there’s also enough information out there to prove that gay people don’t cause earthquakes. People will believe what they want to believe.

§

On the other hand, I’ve said above that I don’t like completely ignoring the accounts of thousands of people who say there’s a problem. Although my female friends in computer science keep insisting they’ve never encountered sexism there, many many others say they have.

But let’s keep our causal arrows pointing the right direction. Any space with a four-to-one male:female ratio is going to end up with some pretty desperate people and a whole lot of unwanted attention. Add into this mix the fact that nerds usually have poor social skills (explaining exactly why would take a literature review to put that last one to shame, but hopefully everyone can agree this is true), and you get people who are pretty sure they are supposed to do something but have no idea what. Err to one side and you get the overly-chivalrous people saying m’lady because it pattern matches to the most courtly and least sexual way of presenting themselves they can think of. Err to the other, and you get people hollowly imitating the behavior they see in famous seducers and playboys, which when done without the very finely-tuned social graces and body-language-reading-ability of famous seducers and playboys is pretty much just “being extremely creepy”.

But once you accept this model, it starts to look like feminists and I are trying to solve the same problem.

The problem is that nerds are scared and confused and feel lonely and have no idea how to approach women. From this root problem blossoms both Aaronson’s problem – that sometimes all you can do is go to a psychiatrist and ask to be castrated – and Penny’s problem – that other times people go read pickup artistry books that promise to tell them how the secret is “negging” people.

But Aaronson’s solution to the problem is to talk about it. And feminism’s solution to the problem is to swarm anyone who talks about it, beat them into submission, and tell them, in the words of Marcotte, that they are “yalping entitlement combined with an aggressive unwillingness to accept that women are human beings just like men”

IX.

Every article about male nerds calls us “entitled”.

I’m pretty sure they don’t mean financially, since nerds for example give disproportionately more to charity than other groups (see: Bill Gates, the joke in the effective altruist movement that it contains “all kinds of people – mathematicians, economists, philosophers, and computer scientists”).

And I’m pretty sure they don’t mean politically, since nerds are far more likely to support wealth redistribution than the general population (compare political alignment here to your choice of nationwide poll).

And I’m pretty sure they don’t mean psychologically. In psychology, entitlement as a construct is usually blended with narcissism. Predictors of narcissism include high emotional intelligence, high social skills but (uniquely among Dark Triad traits) not high nonverbal (ie mathematical) intelligence, and high extraversion. Another interesting fact about narcissists is that they tend to have more sexual partners than non-narcissists. Jonason describes the research on narcissism and sex by saying that “Narcissists find it easy to start new relationships but are less committed to and interested in staying in existing relationships.” I feel like even feminists should be able to agree that “extraverted people with excellent social skills but no particular mathematical aptitude who find it easy to start new relationships” is not a perfect match for nerds here.

So I don’t think these articles are talking about entitlement full stop. I guess they’re using this to point solely at sexual entitlement. But even this seems to require further clarification.

Do they mean nerds hold sexist attitudes? The research (1, 2, 3, 4) shows that sexist attitudes are best predicted by low levels of education, high levels of religious belief, and (whites only) low neuroticism. Once again, I don’t feel it should be controversial to say that “very religious people who drop out of school early and are psychologically completely healthy” is not how most people would describe nerds. Besides, in a survey I did of 1500 people on an incredibly nerdy forum last year, the average was extremely feminist, so much so that the average nerdy man was more feminist than the average non-nerdy woman.

Do they mean nerds are more likely to rape people? There is an appropriate caveat here that it is difficult-to-impossible to profile rapists – but if people took that caveat seriously then you couldn’t profile nerds as rapists either. Since we’re already talking about profiling, let’s go all the way and find that the best research about rapists (source: David Lisak) does find various characteristics of undetected campus rapists (ie primarily date rapists who get away with it, we’re not just talking about scary felons with knives here as a red herring). Some of these are purely psychological (“they’re sexist and don’t like women”). But the rest include: rapists are more sexually active and “engage in consensual and coercive sex far more often than is typical for men of their age group”. They are members of “sexually violent subcultures” including “fraternities and gangs”. They are “hypermasculine” and “strive always to behave in rigidly and stereotypically masculine ways” They are heavy drinkers, often using alcohol to release either their own inhibitions or those of their victims.

Once again, I feel like “hypermasculine frat boys and gangsters who party too hard and have a large number of partners” is a really poor description of nerds.

When people talk about nerds feeling “sexually entitled”, it’s never about any of these things. It’s always the same: A male nerd has dared to express that he is sad about being alone and miserable. Then they round this off to “therefore he believes everyone else owes him sex because he is so great” in precisely the way Amanda Marcotte does explicitly and Penny allows to lie beneath the surface.

Once again, Scott Aaronson’s entire problem was that he was so unwilling to hurt women even unintentionally, and so unclear about what the rules were for hurting women, that he erred on the side of super-ultra-caution and tried to force himself never to have any sexual interest in women at all even to the point of trying to get himself castrated. If entitlement means “I don’t care about women’s feelings, I just care about my own need for sex”, Aaronson is the perfect one hundred eighty degree opposite of entitlement. He is just about the most unentitled (untitled?) person imaginable.

Yet Aaronson is the example upon which these columnists have decided their case for “nerd entitlement” must rise and fall. You have better examples? Then why didn’t you use them?

I’ve already admitted that when a girl asked me out in middle school, I ran away terrified because I figured nobody could actually like me and it was obviously some kind of nasty trick. If entitlement means “believing you deserve all the sex”, then teenage-me also sounds pretty untitled.

Yet I, too, get to forever read articles about how entitled I am.

I’m not making some kind of #NotAllNerds statement here, any more than someone who disagrees with the claim “elephants are tiny” is claiming #NotAllElephants

A better word for this untitlement is, perhaps, scrupulosity, where you believe you are uniquely terrible and deserve nothing. Scrupulosity is often linked to obsessive compulsive disorder, which the recent survey suggests nerds have at higher rates than the general population and which is known to be more common in high-IQ people. When I hear my utilitarian friends say things like “I have money and people starving in Africa don’t have money, therefore I am morally obligated to give half of my money to people starving in Africa or else their starvation is my fault” and then actually go and do that – and trust me, these people are always nerds – then as often as not it’s scrupulosity at work.

When you tell a highly-untitled, high-scrupulosity person that they are entitled, it goes about as well as telling an anorexic person that they are fat.

If your excuse is going to be “okay, some nerds are overly scrupulous, but others are entitled”, how come that wasn’t your argument before? And how come, with laser-like focus, you only pick on the scrupulous ones? How come it’s 2015 and we still can’t agree that it’s not okay to take a group who’s already being bullied and harassed, stereotype it based on the characteristics of its worst members, and then write sweeping articles declaring that the entire group is like that?

X.

When Laurie Penny writes to women, she says:

What I most wanted to say, to all the messed-up teenagers and angry adults out there, is that the fight for your survival is political. The fight to own your emotions, your rage and pain and lust and fear, all those unspeakable secrets that we do not share because we worry that we will be hurt or shunned, is deeply political.

When Laurie Penny writes to men, she says:

Most of all, we’re going to have to make like Princess Elsa and let it go – all that resentment. All that rage and entitlement and hurt.

Clearly this second suggestion contains a non-standard use of the word “we”.

When women feel like they’re not allowed to “own their emotions” like “lust”, or have “secrets that they do not share because they worry that they will be hurt or shunned”, then it is “deeply political” and they have to “fight about it.”

When men make the same complaint, they are encouraged to “let go” of their “resentment” and “entitlement”.

The same worries, deep and secret fears, that are the core and driving heat of Penny’s feminism when they happen to women get called “entitlement” when they happen in men and need to be “let go”. You’re not allowed to complain about them. You’re not even allowed to ask the people hurting you to stop – then you’re super entitled. You shut up and get on with your life.

But it’s actually much worse than that. If you remember only one thing from this entire post, remember that Anakin Skywalker is unbearably annoying remember this:

The past is over. I do not hold, and have never held, any ill will toward the women who rejected me. Some of them continue to be my close friends. Some of them I’ve talked to about this Scott Aaronson thing, and even they agree with me on it. Nor did Aaronson mention any ill will to anyone who rejected him. Talking about how nerds should let go of our past resentment to our crushes is a giant red herring.

What this entire discussion is about is our very present resentment toward the (some) feminists who continue to perpetuate the stereotypes that hurt us then, continue to attack us now whenever we talk about the experience or ask them to stop, and continue to come up with rationalizations for why they don’t have to stop. This isn’t about little Caitlin who wouldn’t return my eye contact in seventh grade, this is about Amanda Marcotte, Jezebel, Gawker, and an entire system that gets its jollies by mocking us and trying to twist the knife.

The only reason little Caitlin is being brought up is so that feminists who don’t want to stop twisting can sidestep any criticism by pretending our argument is entirely how a seventh-grader shouldn’t have control of her own romantic decisions.

@#!$ that. Little Caitlin can do what she wants with her life. But dehumanizing and perpetrating stereotypes about a whole group of people who already have it pretty bad is not okay.

XI.

I already know that there are people reading this planning to write responses with titles like “Entitled Blogger Says All Women Exist For His Personal Sexual Pleasure, Also Men Are More Oppressed Than Women, Also Nerds Are More Oppressed Than WWII Era Jews”. And this post is way too long for most people who read those responses to get their misconceptions corrected. So before I close, let me give a brief summary of what I am trying to say:

1. There are a lot of really nasty stereotypes perpetuated about nerds, especially regarding how they are monsters, nobody can love them, and they are too disgusting to have relationships the same way other people do.

2. Although both men and women suffer from these stereotypes, men really do have a harder time getting relationships, and the experience is not the same.

3. Many of the people suffering from these stereotypes are in agreement that it is often self-identified feminists who push them most ardently, and that a small but vocal contingent of feminists seem to take special delight in making nerds’ lives worse.

4. You cannot define this problem away with the word “patriarchy”.

5. You cannot define this problem away by saying that because Mark Zuckerberg is a billionnaire, nerds are privileged, so they already have it too good. The Jews are a classic example of a group that were both economically advantaged in a particular industry, but also faced unfair stereotypes.

6. Whether women also have problems, and whether their problems are even worse, is not the point under discussion and is not relevant. Women can have a bunch of problems, but that doesn’t mean it is okay for any feminists to shame and bully nerds.

7. Nerds are not uniquely evil, they are not especially engaged in oppressing women, and they are not driving women out of Silicon Valley. Even if they were, “whenever they choose to open up about their private suffering” is not the time to talk about these things.

8. “Entitlement” is a uniquely bizarre insult to level at nerds given that by most of the term’s usual definitions nerds are some of the most untitled people there are.

9. The feminist problem of nerds being desperate and not having any social skills (and therefore being creeps to women) is the same as the nerd problem of nerds being desperate and not having any social skills (and therefore having to live their life desperate and without social skills). Denying the problem and yelling at nerds who talk about it doesn’t help either group.

10. The nerd complaint on this issue is not “high school girls rejected us in the past when we were lonely and desperate,” it is “some feminists are shaming us about our loneliness and desperation in the past and present and openly discussing how they plan to do so in the future.” Nobody with principles is angry at the girls who rejected them in the past and this is a giant red herring. If you don’t believe any feminists are shaming anyone, then say so; don’t make it about little Caitlin in seventh-grade.

If you want to debate or fisk this article, I would recommend using these paragraphs as starting points instead of whatever bizarre perversions of my words the brain of the worst person reading this can dream up.

[EDIT 1/15: Okay, it looks like the talking point people chose to go with was “he made a 1984 joke, therefore the thesis of the essay is that all men are oppressed by all women exactly as badly as people are oppressed in 1984.” As usual, I was insufficiently pessimistic.]

XII.

Penny ends:

We bring our broken hearts and blue balls to the table when we talk gender politics, especially if we are straight folks. Consent and the boundaries of consent – desire and what we’re allowed to speak of desire – we’re going to have to get better, braver and more honest, we’re going to have to undo decades of toxic socialisation and learn to speak to each other as human beings in double quick time.

[…]

The road ahead will be long. I believe in you. I believe in all of us. Nerds are brilliant. We are great at learning stuff. We can do anything we put our minds to, although I suspect this thing, this refusing to let the trauma of nerdolescence create more violence, this will be hardest of all.

I see a vision here of everybody, nerdy men, nerdy women, feminists, the media, whoever – cooperating to solve our mutual problems and treat each other with respect. Of course I am on board with this vision. As Scott Aaronson would put it, I am 97% on board. What keeps me from being 100% on board right now is the feeling that the other side still doesn’t get it.

First of all, a whole lot of other side is not Laurie Penny. They are the people gleefully mocking our pain and telling us we deserve it. But even the good people are worrisome enough.

They admit that nerdy men, lesbians, bisexuals, etc may be in pain, but they deny categorically any possible role of feminist shaming culture in causing that pain and want to take any self-reflection on their part off of the table of potential compromise.

They admit that our pain technically exists, but they are unable to acknowledge it without adding “…but by the way, your pain can’t possibly ever be as bad as our pain” or “your pain doesn’t qualify for this ontologically distinct category of pain which is much more important.”

They continue to think it is appropriate to respond to any complaint or expression of suffering on our part with accusations of “entitlement”, comparisons to Darth Vader, and empirically-contradicted slanders about how our mere presence drives women away from everything we love.

Once I see anyone, anywhere, publish an article that not only recognizes our pain, but doesn’t derail it into an explanation of why we’re definitely still terrible and there is no need whatsoever for them to change, then I will be more optimistic that progress is at hand.

XIII.

Oh frick.

And on that note I shall return to what I was doing before I read this post, which was drinking sweet tea and weeping about how boys don’t seem to want to kiss short-haired lady nerds, and trying not to blame the whole world for my broken heart, which is becoming more complex and interesting in the healing but still stings like a boiling ball of papercuts. I’ll let you know how that goes.

Having so much fun picking this article apart, and then this 🙁

Look. I mean what I say about how I don’t believe in zero-sum games. The reality of Prof. Aaronson’s problem does not for one second diminish the reality of Ms. Penny’s sadness as well.

So here is my offer to Ms. Penny. If she accepts and is in some kind of heavily nerd-populated city (NYC? SF?) I will use my connections in the nerd community to get her ten dates within ten days with intelligent, kind, respectful nerdy men of whom she approves. If she is in some less populated place, I will get her some lesser but still non-zero number of dates (unless she’s in Greenland or somewhere, in which case she’s on her own).

If I can’t do that, she may feel welcome to publically mock me and tell me that I was overconfident about how many people are, in fact, extremely willing to kiss short-haired lady nerds.

The rest of this article was serious, but this is extra serious. Let me know.

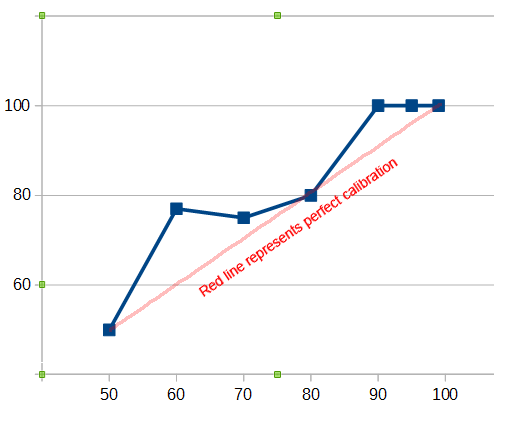
[EDIT: Comments are now closed, because this got linked on Instapundit and I know from experience that bad things happen if you leave the comments open after that point. Also, my comment software starts acting weird after like a thousand. If you must comment on this further, go bother Ozy on their open thread]. If you’re named in this article and you want to rebut it or reply, email me and I’ll include it somewhere.

# 2014 Predictions: Calibration Results

Last year in January I made some predictions for 2014 and gave my calibration level for each. Time to see how I did. Ones in black were successes, ones in red were failures.

1. Obamacare will survive the year mostly intact: 80%  
2. US does not get involved in any new major war with death toll of > 100 US soldiers: 90%  
3. Syria’s civil war will not end this year (unstable cease-fire doesn’t count as “end”): 60%  
4. Bitcoin will end the year higher than $1000: 70%  
5. US official unemployment rate will be < 7% in Dec 2014: 80%  
6. Republicans will keep the House in US midterm elections: 80%  
7. Democrats will keep the Senate in US midterm elections: 60%  
8. North Korea’s government will survive the year without large civil war/revolt: 80%  
9. Iraq’s government will survive the year without large civil war/revolt: 60%  
10. China’s government will survive the year without large civil war/revolt: 99%  
11. US government will survive the year without large civil war/revolt: 99%  
12. Egypt’s government will survive the year without large civil war/revolt: 60%  
13. Israel-Palestine negotiations remain blocked, no obvious plan for Palestinian independence: 99%  
14. Israel does not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 95%  
15. Sochi Olympics will not be obvious disaster (ie spectacular violence, protests that utterly overshadow events, or have to be stopped early): 90%  
16. Putin will remain President of Russia at end of 2014: 95%  
17. Obama will remain President of USA at end of 2014: 95%  
18. No nuclear weapon used in anger in 2014: 99%  
19. No terrorist attack in USA killing > 100 people: 90%  
20. No mass shooting in USA killing > 10 people: 50%  
21. Republic of Shireroth will not officially disband or obviously die before end of 2014: 90%  
22. Republic of Shireroth will remain in Bastion Union: 80%  
23. Active population of Shireroth on last 2014 census will be between 10 and 20: 70%  
24. Slate Star Codex will remain active until end of 2014: 70%  
25. Slate Star Codex will get more hits in 2014 than in 2013: 60%  
26. At least one 2014 Slate Star Codex post will get > 10,000 hits total: 80%  
27. No 2014 Slate Star Codex post will get > 100,000 hits total: 90%  
28. 2014 Less Wrong Survey will show higher population than 2013 Survey conditional on similar methodology: 80%  
29. 2014 Less Wrong Survey will show population < 2000 C.O.S.M.: 70%  
30. 2014 Less Wrong Survey will have > % female than 2013 Less Wrong Survey: 70%  
31. 2014 Less Wrong Survey will have < 20% female: 90%  
32. HPMoR will conclude in 2014: 80%  
33. At least 1 LW post > 100 karma in 2014: 50%  
34. No LW post > 100 karma by me: 80%  
35. CFAR will continue operating in end of 2014: 90%  
36. MIRI will continue operating in end of 2014: 99%  
37. MetaMed will continue operating in end of 2014: 80%  
38. None of Eliezer, Luke, Anna, or Julia will quit their respective organizations: 60%  
39. No one in LW community will become world-famous (let’s say >= Peter Thiel) for anything they accomplish this year: 80%  
40. MIRI will not announce it is actively working on coding a Friendly AI (not just a few bits and pieces thereof) before the end of 2014: 99%  
41. I will remain at my same job through the end of 2014: 95%  
42. I will get a score at >95th percentile for my year on PRITE: 70%  
43. I will be involved in at least one published/accepted-to-publish research paper by the end of 2014: 20%  
44. I will not break up with any of my current girlfriends through the end of 2014: 50%  
45. I will not get any new girlfriends in 2014: 50%  
46. I will not be engaged by the end of 2014: 80%  
47. I will be living with Ozy by the end of 2014: 80%  
48. I will take nootropics on average at least once/week through the second half of 2014: 50%  
49. I will not manage to meditate at least 100 days in 2014: 80%  
50. I will attend NYC Solstice ritual: 60%  
51. I will arrange some kind of Michigan Solstice Ritual: 50%  
52. I will not publicly identify as religious (> atheist) by the end of 2014: 95%  
53. I will not publicly identify as neoreactionary or conservative (> liberal or libertarian) by the end of 2014: 70%  
54. I will not publicly identify as leftist or communist (> liberal or libertarian) by the end of 2014: 80%  
55. I will get a Tumblr in 2014: 50%  
56. I will not delete/abandon either my Facebook or Twitter accounts: 60%  
57. I will have less than 1000 Twitter followers by the end of 2014: 60%  
58. When Eliezer sends me a copy of “Perfect Health Diet”, I will not be convinced that it is more correct or useful than the best mainstream nutrition advice (eg Stephen Guyenet’s blog): 70%  
59. I will end up being underconfident on these predictions: 50%

Of predictions at the 50% level, 4/8 (50%) were correct  
Of predictions at the 60% level, 7/9 (77%) were correct  
Of predictions at the 70% level, 6/8 (75%) were correct  
Of predictions at the 80% level, 12/15 (80%) were correct  
Of predictions at the 90% level, 7-0 (100%) were correct  
Of predictions at the 95% level, 5-0 (100%) were correct  
Of predictions at the 99% level, 6-0 (100%) were correct



I declare myself to be impressively well-calibrated. You should all trust me about everything.

2015 predictions coming soon.

# Ethics Offsets

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.

# Chronic Psychitis

Some people have asked my opinion on a recent spate of articles like Is Depression Partly Caused By An Allergic Reaction? and Depression May Be Caused By Inflammation.

Standard disclaimer: I’m not a researcher in this field, I’m not board-certified as a full psychiatrist yet, and what I remember of biochemistry is limited to being pretty sure there’s something called a “Krebs cycle” involved somewhere. That having been said:

This is pretty legit.

Start with From inflammation to sickness and depression, Dantzer et al (2008), who note that being sick makes you feel lousy [citation needed]. Drawing upon evolutionary psychology, they theorize this is an adaptive response to make sick people stay in bed (or cave, or wherever) so the body can focus all of its energy on healing. A lot of sickness behavior – being tired, not wanting to do anything, not eating, not wanting to hang around other people – seems kind of like mini-depression.

All of this stuff is regulated by chemicals called cytokines, which are released by immune cells that have noticed an injury or infection or something. They are often compared to a body-wide “red alert” sending the message “sickness detected, everyone to battle stations”. This response is closely linked to the idea of “inflammation”, the classic example of which is the locally infected area that has turned red and puffy. Most inflammatory cytokines handle the immune response directly, but a few of them – especially interleukin-1B and tumor necrosis factor alpha – cause this depression-like sickness behavior. It is noted that:

In general, animals injected with IL-1ß or TNF-a stay in a corner of their home cage in a hunched posture and show little or no interest in their physical and social environment unless they are stimulated. Specifically, they show decreased motor activity, social withdrawal, reduced food and water intake, increased slow-wave sleep and altered cognition

Here are some other suspicious facts about depression and inflammation:

– Exercise, good diet and sleep reduce inflammation; they also help depression.

– Stress increases inflammation and is a known trigger for depression.

– Rates of depression are increasing over time, with the condition seemingly very rare in pre-modern non-Westernized societies. This is commonly attributed to the atomization and hectic pace of modern life. But levels of inflammation are also increasing over time, probably because we have a terrible diet that disrupts the gut microbiota that are supposed to be symbioting with the immune system. Could this be another one of the things we think are social that turn out to be biological?

– SSRI antidepressants, like most medications, have about five zillion effects. One of the effects is to reduce the level of inflammatory cytokines in the body. Is it possible that this is why they work, and all of this stuff about serotonin receptors in the brain is a gigantic red herring?

– It’s always been a very curious piece of trivia that treating depression comorbid with heart disease significantly decreases your chances of dying from the heart disease. People just sort of nod their heads and say “You know, mind-body connection”. But inflammation is known to be implicated in cardiovascular disease. If treating depression is a form of lowering inflammation, this would make perfect sense.

– Rates of depression are much higher in sick people. Cancer patients are especially famous for this. No one gets too surprised here, because having cancer is hella depressing. But it’s always been interesting (to me at least) that as far as we can tell, antidepressants treat cancer-induced depression just as well as any other type. Are antidepressants just that good? Or is the link between cancer being sad and cancer causing depression only part of the story, with the other part being that the body’s immune response to cancer causes inflammatory cytokine release, which antidepressants can help manage?

– Along with cancer, depression is common in many other less immediately emotion-provoking illnesses like rheumatoid arthritis and diabetes. The common thread among these illnesses is inflammation.

– Inflammation changes the activity level of the enzyme indoleamine 2,3 dioxygenase. This enzyme produces kynurenines which interact with the NMDA receptor, a neurotransmitter receptor implicated in depression and various other psychiatric diseases (in case your first question upon learning about this pathway is the same as mine: yes, kynurenines got their name because they were first found in dog urine).

– Sometimes doctors treat diseases like hepatitis by injecting artificial cytokines to make the immune system realize the threat and ramp up into action. Cytokine administration treatments very commonly cause depression as a side effect. This depression can be treated with standard antidepressants.

– Also, it turns out we can just check and people with depression have more cytokines.

There’s also some evidence against the theory. People with depression have more cytokines, but it’s one of those wishy-washy “Well, if you get a large enough sample size, you’ll see a trend” style relationships, rather than “this one weird trick lets you infallibly produce depression”.

But for me the strongest evidence against is a general feeling that it’s very easy to get lots of convincing evidence for a theory in medicine whether or not it’s true.

Twenty years ago, everyone was super-convinced that depression was caused by low serotonin levels. We found that depressed people on average had lower serotonin levels than non-depressed people. We found that giving people drugs that increased serotonin treated depression. We did lots of studies proving serotonin was a vital chemical that regulated mood. We found that genes affecting serotonin-related proteins were linked to depression. We did PET scans that found abnormally high levels of activity in serotonin-related enzymes in the brains of depressed people. It was all very convincing. And right now everyone’s pretty sure it’s wrong.

Ten years ago, everyone was super-convinced that depression was caused by under-secretion of the neuro-hormone BDNF and subsequent decline in hippocampal neurogenesis. It was dutifully found that depressed people had less BDNF than everyone else, and less hippocampal neurogenesis. Exercise, sleep, good diet, and all the other things that help depression were found to also raise levels of BDNF. Chemical pathways were trotted out by which effective antidepressants would probably raise BDNF levels. I think this theory is still very popular, but for the inflammation theory to be right someone will either have to disprove this one or tie it together with some theory of why inflammation decreases BDNF or low BDNF increases inflammation or something else. I do see some evidence that this is true, but to fully integrate the theories is going to take a lot more than that.

And these are just the two most recent and most famous. We have Freud’s psychoanalytic theory of depression, lots of people studying dysregulation of the hypothalamopituitaryadrenocortical axis, some pointers to dysregulation in the second messenger system, et cetera. All of these theories have great evidence.

Point is, now we have another theory that neatly explains how depression starts, how antidepressants work, why diet and exercise are good for you, and all the things all the other theories explained. Maybe the third time’s the charm?

A lot of the things in the body are really complex. Inflammation definitely affects serotonin – the indoleamine 2,3 dioxygenase enzyme acts on serotonin’s immediate precursor. It affects BDNF levels, as above, which in turn affect hippocampal neurogenesis. The hypothalamopituitaryadrenocortical axis releases cortisol, which downregulates the immune system and decreases the action of inflammatory cytokines. All of the anxiety-inducing life events and intrapsychic conflicts and secret desires to marry your mother that Freud thought caused depression produce a lot of stress, which both releases cortisol and reduces normal ability to regulate inflammatory response.

So basically all of these systems are intimately interconnected, and probably before this is done with researchers will find five more systems intimately interconnected with all of these. It might be that inflammation is the master system which causes a cascade of events in all of the others. It might be that one of the others is the master system. It might be that depression is a collection of multiple different diseases, and some are caused by one thing and others by another. It might be that looking for a “master system” is silly and that the true mathematical relationship between all of these things is such a chaotic process that all you can say is that they all stumbled together into the wrong attractor point and things deteriorated from there.

Anyway, all this is for much smarter people than me to figure out. The question I’m most interested in: can we treat depression by giving people anti-inflammatory drugs?

The answer seems to be: it depends how strongly you object to getting a heart attack.

Aspirin is a great anti-inflammatory drug. It’s pretty safe in adults (except for a small risk of GI bleeding) and it decreases risk of cardiovascular disease and cancer as well. If you could treat depression with aspirin, you’d be home free. However, the most convincing review I have seen for aspirin is unimpressed. It points out that some trials have shown negative effects for aspirin, and that long-term use of aspirin can increase intestinal permeability which decreases ability to regulation inflammation which is the opposite of what we want. Right now there isn’t much evidence on this issue, but what there is isn’t promising.

Most researchers have chosen to focus on celecoxib (Celebrex™®©, a high-tech next-generation anti-inflammatory). Here the evidence is actually very strong. Last month’s JAMA Psychiatry contained Effect of anti-inflammatory treatment on depression, depressive symptoms, and adverse effects: a systematic review and meta-analysis of randomized clinical trials, Kohler et al, (2014), which analyzes ten studies with a total of 4000 people taking celecoxib and finds an effect size similar to that of SSRI antidepressants. This is promising and exciting.

They add: “We found no evidence of an increased number of gastrointestinal or cardiovascular events after 6 weeks.”

That’s probably because they didn’t wait long enough. Celecoxib is very closely related to the infamous rofecoxib (Vioxx™®©) which got pulled from the market for quadrupling heart attack risk. Celecoxib is safer; it only increases your risk by some smaller amount depending on dose. Studies conflict, but maybe 33% for a standard regimen?

On the other hand, if ten percent of Americans are on SSRIs right now, and there are 1.5 million heart attacks per year in the US, and celecoxib increases that by 33%, then switching everyone from SSRI to celecoxib would cause…quick Fermi calculation…ignore interactions…50,000 extra heart attacks per year. Ouch.

Celecoxib is a good drug for its indicated uses, which involve treating chronic pain conditions that nothing else can treat safely. But it’s hardly something I’d want to start giving to every depressed patient who walks into a psychiatrist’s office. Maybe as a third line or fourth line drug for desperate people. But then, we already have plenty of good third-line and fourth-line drugs for desperate people. You want strong psychiatric medication and aren’t too concerned about the state of your cardiovascular system? Here, have an antipsychotic!

So in conclusion, I think the inflammatory hypothesis of depression is very likely part of the picture. Whether it’s the main part of the picture or just somewhere in the background remains to be seen, but for now it looks encouraging. Anti-inflammatory drugs do seem to treat depression, which is a point in the theory’s favor, but right now the only one that has strong evidence behind it has side effects that make it undesirable for most people. There’s a lot of room to hope that in the future researchers will learn more about exactly how this cytokine thing works and be able to design antidepressant drugs that target the appropriate cytokines directly. Until then, your best bets are the anti-inflammatory mainstays: good diet, good sleep, plenty of exercise, low stress levels, and all the other things we already know work.

# The Phatic And The Anti-Inductive

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.

# The Physics Diet?

There are at least four possible positions on the thermodynamics of weight gain:

1. Weight gain does not depend on calories in versus calories out, even in the loosest sense.

2. Weight gain is entirely a function of calories in versus calories out, but calories may move in unexpected ways not linked to the classic “eat” and “exercise” dichotomy. For example, some people may have “fast metabolisms” which burn calories even when they are not exercising. These people may stay very thin even if they eat and exercise as much as much more obese people.

3. Weight gain is entirely a function of calories in versus calories out, and therefore of how much you eat and exercise. However, these are in turn mostly dependent on the set points of a biologically-based drive. For example, some people may have overactive appetites, and feel starving unless they eat an amount of food that will make them fat. Other people will have very strong exercise drives and feel fidgety unless they get enough exercise to keep them very thin. These things can be altered in various ways which cause weight gain or loss, without the subject exerting willpower. For example, sleep may cause weight loss because people who get a good night sleep have decreased appetite and lower levels of appetite-related hormones.

4. Weight gain is entirely a function of calories in versus calories out, and therefore of how much you eat and exercise. That means diet is entirely a function of willpower and any claim that factors other than amount of food eaten and amount of exercise performed can affect weight gain is ipso facto ridiculous. For example, we can dismiss claims that getting a good night’s sleep helps weight loss, because that would violate the laws of thermodynamics.

1 and 4 are kind of dumb. 1 is dumb because…well, to steal an Eddington quote originally supposed apply to the second law of thermodynamics:

If someone points out to you that your pet theory of the universe is in disagreement with Maxwell’s equations — then so much the worse for Maxwell’s equations. If it is found to be contradicted by observation — well, these experimentalists do bungle things sometimes. But if your theory is found to be against…thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation.

But 4 is also dumb. We have a long list of things that affect weight gain – for example, patients on the powerful psychiatric medication clozapine usually gain a lot of weight – fifteen pounds more on average than people on safer antipsychotics. Other medications are known to increase weight to a lesser degree, and some medications even decrease weight, though you wouldn’t like the side effects of most of them. Certain genetic diseases are also known to cause increased weight – Prader-Willi syndrome, for example.

One could try to rescue 4 by saying that people with rare genetic diseases or taking powerful prescription-only medications are a different story and in normal people it’s entirely controlled by willpower. But first, this is an area where possibility proofs are half the battle, and we have a possibility proof. And second, there are more than enough studies about genetics, microbiome, and, yes, sleep showing that all of these things can have effects in normal people.

So 1 and 4 are out. And although I do sometimes see people pushing them, they mostly seem to do a thriving business as straw men for people who want to accuse their opponents of saying something absurd.

The most interesting debate to be had is between 2 and 3. 3 says that all of the interventions that we know affect weight – certain pills, certain recreational drugs, changes in gut bacteria, whatever – do it by affecting appetite and exercise drive. 2 says that basal metabolism is also involved. 3 seems to at least leave open the possibility of just starving yourself even when your body is telling you really hard to eat. 2 says even that won’t work.

There’s room for a little bit of gradation between 2 and 3. A lot of people suggest that one way “fast metabolism” presents is by people fidgeting a lot, which is sort of the same as “your body increases its exercise drive”.

But in general, I think 2 is an important issue that does cause at least some interpersonal weight differences.

We’ll start with the “possibility proof” again. MRAP2. It’s a gene. Scientists can delete it in mice. These mice will eventually develop excessive appetites. But when they are young, they eat the same amount as any other mouse, but still get fatter.

Likewise, 2,4-dinitrophenol is a cellular uncoupling agent which increases metabolic rate and consistently produces weight loss of 2-3 pounds per week. It would be an excellent solution to all of our obesity-related problems if the papers on it didn’t keep having names like 2,4-Dinitrophenol: A Weight Loss Agent With Significant Acute Toxicity And Risk Of Death.

So what about everyday life?

A study of individual variation in basal metabolic rate found very significant interpersonal differences. A lot of that was just “some people are bigger than others”, but some of it wasn’t – they state that “twenty-six percent of the variance remained unexplained”. The Wikipedia article puts this in context: “One study reported an extreme case where two individuals with the same lean body mass of 43 kg had BMRs of 1075 kcal/day (4.5 MJ/day) and 1790 kcal/day (7.5 MJ/day). This difference of 715 kcal/day (67%) is equivalent to one of the individuals completing a 10 kilometer run every day”

Dr. Claude Bouchard and his team stuck 12 pairs of male identical twins in isolation chambers where their caloric intake and exercise could be carefully controlled, then fed them more calories than their bodies needed. All sets of twins gained weight, and in all twin groups both twins gained about the same amount of weight as each other, but the amount of weight gained varied between twin pairs by a factor of 3 (from 4 to 13 kg).

A lot of the sites that talk about this thing are careful to say that people “can’t blame” genes for their obesity, because obesity levels have been rising for decades and genes can’t change that quickly. I think this is wrong-headed. True, genes are not the source of the modern rise in obesity levels. But it’s entirely possible that a globally rising tide of obesity has disproportionately affected the people with the wrong genes. Just as Bouchard fed the same amount extra to all his study participants but some of them gained more weight than others, so if you put an entire civilization worth of people in an obesogenic environment, some of them might be genetically predisposed to do worse than the rest.

A more practical question – can individual people’s metabolism change?

I am personally predisposed to answer in the affirmative. In my early twenties, I ate a crazy amount every day – two bagels with breakfast, cookies with lunch, a big dinner followed by dessert – and I stayed pretty thin throughout. Now I’m thirty, I eat a very restrained diet, and my weight still hovers at just above the range where I am supposed to be. I know that people are famously bad at understanding how much they’re eating and exercising, but seriously if you try to convince me that I’m eating more now than I was then I’m going to start doubting my own sanity, or at least my autobiographical memory.

But there’s not much evidence to back me up. Metabolic rate is well-known to decline with age, but linearly and predictably. And it changes with muscle mass, but only minimally – and I don’t think I used to be any more muscular.

The sites that talk about drastic and unexpected ways to change metabolism seem mostly crackpottish. This isn’t to say their methods don’t work – green tea, for example, has a statistically significant effect – but it’s all so small as to be pretty meaningless in a real-world context.

So my own story seems to be on shaky ground. But as far as I can tell, the people arguing that they’re trying just as hard as anybody else but still unable to lose weight because of their metabolism are very possibly right.

# The Influenza Of Evil

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.

# Depression Is Not A Proxy For Social Dysfunction

I.

Here is a terrible article from the New York Post: Sorry, Liberals, Scandinavian Countries Aren’t Utopias.

Its thesis is interesting and worth exploring, but instead of a principled investigation, the article just publishes a bunch of cherry-picked smears about Scandinavia. Did you know that 5% of Danes have had sex with animals?

(What percent of people in other countries have had sex with animals? I don’t know. More important, I see no sign that the New York Post knows either.)

But the part that really caught my eye was statements like these:

Why does no one seem particularly interested in visiting Denmark? Visitors say Danes are joyless to be around. Denmark suffers from high rates of alcoholism. In its use of antidepressants it ranks fourth in the world. (Its fellow Nordics the Icelanders are in front by a wide margin) … Finland, which tops the charts in many surveys, is also a leader in categories like alcoholism, murder, suicide and antidepressant usage.

The Post is not the only paper to make this argument. The Guardian (“The Grim Truth Behind The Scandinavian Miracle”) has said much the same thing:

Take the Danes, for instance. True, they claim to be the happiest people in the world, but why no mention of the fact they are second only to Iceland when it comes to consuming anti-depressants?…Finland has by far the highest suicide rate in the Nordic countries.

I’ve heard this same argument applied to other issues; for example, in his debate with Noah Smith, Michael Anissimov argues against the supposed success of modern liberal society by pointing out rising rates of depression and suicide.

It’s really tempting to equate depression with misery and misery with social dysfunction. Danes and Finns have high levels of depression, therefore their lives must be unusually miserable, therefore Denmark and Finland are poorly-organized societies.

But first of all, it’s not clear that Scandinavian countries really have very high depression and suicide rates. There are a lot of collections of statistics, and many of them show Scandinavia around the middle. Going by “antidepressant prescriptions” is a terrible way to do things, because it mixes amount of depression with resources devoted to treating depression – if the Scandinavian health systems are as good as everyone says, maybe they just treat a greater percent of their depressives than everywhere else.

But more important, even if Scandinavia does have very high rates of depression, that doesn’t tell us much about whether they’re happy or not. Depression is not the same thing as being sad. Sadness is a risk factor for depression – although even there I suspect that it’s very specific kinds of sadness that we haven’t yet teased out from the general construct – but it is not the condition itself. The condition itself is a complicated mess of neurotransmitters, cytokines, hormones, changes in brain structure, and goodness only knows what else.

Off the top of my head, here are six plausible reasons why Scandinavia could have higher rates of depression than the United States, even if it is a utopian society of perfect happiness.

1. Light. Scandinavia is far north [citation needed] which puts its citizens at very high risk for seasonal affective disorder, which can present as depression.

2. The midnight sun. Scandinavia’s weird day-night cycle could easily disrupt people’s circadian rhythms. Studies find that “increasing evidence points to a role of the biological clock in the development of depression…it seems likely the circadian system plays a vital role in the genesis of the disorder. This is why some European countries use melatonergic substances as antidepressants.

3. Parasite load. It’s positively correlated with temperature, which means Scandinavia probably has some of the lowest parasite load in the world. But low parasite load causes the immune system to get antsy and start attacking random stuff, leading to increase risk of autoimmune disease. If there’s an immunological component to depression – and right now lots of people think there is – then that’s another risk factor right there.

4. Diet. The Scandinavian diet has unusually little fresh food, because the area is a frozen wasteland and most things have to be imported from elsewhere. They’re big on frozen stuff, processed stuff, and canned stuff. I am neither an expert in Scandinavian cuisine nor in nutrition, but if depression is linked to diet and imbalance in the gut microbiome, which there’s some evidence it is, then diet is heavily implicated and the Scandinavians are in a good position to get hit extra hard.

5. Genetics. The New York Post article mentions that Scandinavians have an unusual variant of the MAO-A enzyme (I told you it was a weird hit piece. Scandinavia is too liberal, therefore they have bad genes?). MAO-A is also known as “the thing that processes serotonin” and “the thing that MAO inhibitors, some of the most powerful known antidepressants, inhibit”. I’m not saying this gene in particular is responsible for Scandinavian depression, I’m saying that the article itself is admitting that Scandinavia contains some genetically distinct populations and for all we know this could be involved.

6. Culture. Maybe the biggest factor in the level of depression and suicide in a culture is whether it is culturally acceptable to be depressed and commit suicide. Some of the lowest suicide rates are found in heavily religious cultures and communities who believe suicide is a mortal sin. On the other hand, one of the most suicidal countries in the world is Japan, with its heavily-mythologized history of heroic samurai taking “the honorable way out” when they had brought shame upon themselves. Well, Scandinavia is one of the least religious regions in the world. And all I know about their culture is that they produce about 100% of good death metal, and their native mythology ends with the world being plunged into eternal winter and the gods being eaten by wolves.

II.

But all this is just speculation. Let me give a concrete example of a case where social dysfunction doesn’t track depression and suicidality in a predictable way.

What about white versus black Americans? To some degree these two groups live in separate “societies”. Most people would consider the white society better off in most ways – higher income, better health, more family stability, less involvement with the criminal justice system. If White America and Black America were countries, White America would get all of the accolades currently given to the Scandinavians.

But American whites have higher rates of depression than blacks. There are the usual contradictory studies and arguments about how to adjust for which confounder, but I’m pretty sure this is something like a consensus position right now. More solidly, white Americans have much higher suicide rates than black Americans.

(although I feel bad mentioning this, because the stereotype that blacks never commit suicide is wrong and sometimes prevents black people from getting the help they need.)

We can go a few centuries back and get even more surprising results. Although it’s difficult to get data from the era, analyses of suicide rate among African-American slaves in the antebellum South describe it as “surprisingly low”. I can’t find any hard evidence proving Kurt Vonnegut’s contention that “the suicide rate per capita among slave owners was much higher than the suicide rate among slaves”, but it seems to have been commonly believed. Kneeland writes:

“[These low suicide rates are] consistent with suicide rates for Africa and for people of African descent living in other areas of the world, and further supports the theory that a low suicide rate is an element of African culture.”

If you’re going to say that Scandinavia’s higher depression and suicide rates mean Scandinavia has it worse off than America, you also need to theorize that white people have it worse off than black people, including black slaves. Why don’t you go post something to that effect on Tumblr and see what they have to say? I’ll wait.

III.

Or maybe we’re barking up entirely the wrong tree. What if it’s not even that happy, well-functioning societies can sometimes still end up with high suicide rates? What if people become suicidally depressed precisely because they live in happy, well-functioning societies?

This is the fascinating hypothesis of Daly, Oswald, and Wu (2011), who after crunching the numbers find pretty convincingly that “suicide rates tend to be highest in happy places”:

A little-noted puzzle is that many of [the happiest] places have unusually high rates of suicide. While this fact has been remarked on occasionally for individual nations, especially for the case of Denmark, it has usually been attributed in an anecdotal way to idiosyncratic features of the location in question (eg the dark winters in Scandinavia), definitional variations in the measurement of well-being and suicide, and differences in culture and social attitudes regarding happiness and taking one’s life. Most scholars have not thought of the anecdotal observation as a systematic relationship that might be robust to replication or investigation…this paper attempts to document the existence of a happiness-suicide paradox: happier areas have a higher percentage of suicides.

They then go on to show a strong positive relationship between average self-reported happiness and suicidality across Western nations – Greece is both the least happy country and the one with the lowest suicide rate – and US states, where confirmed hellholes New York and New Jersey are at or near the bottom. The relationship holds whether you adjust for confounders (including income!) or not.

I expected this to be a straightforward effect of modernization/industrialization/liberalism, as per Michael Anissimov’s hypothesis. The country-level data maybe sort of vaguely supports that trend – Greece and Portugal are our token incompletely-modernized countries and have very low suicide rates, Scandinavia is high, and everywhere else is sort of a toss-up. But US states really really don’t support that hypothesis – New York and Jersey both seem high on the modernization/industrialization/liberalism axis, and they’re right in the bottom left corner of the study’s graphs along with Greece and Portugal. Meanwhile, tropical paradise Hawai’i is suicidal as heck, even though it doesn’t seem espcially modern/industrial/liberalized. The US state data also torpedo – albeit less conclusively – an attempt to make the whole issue one of latitude.

One caveat I do have about the US data is that several of the happiest and most suicidal states – at least on the unadjusted plot – are also high-altitude. Utah, Wyoming, Colorado, Montana, Idaho are all up there at the top left side of the graph. But we already know there’s a strong positive relationship between altitude and suicide in 2584 US counties, probably because the brain’s emotional regulation system doesn’t work well in low-oxygen environments. If we assume people living in beautiful open forested mountain areas are especially happy, that takes away a big chunk of the graph right there. But it leaves other chunks untouched, and I don’t think it’s going to be that simple.

The authors’ preferred explanation is that suicide is an effect of relative rather than absolute misery. If you’re depressed and everybody around you is very happy, that makes things worse than if you’re depressed and everyone around you is also pretty miserable. Thus suicide is more common in happier societies.

I really don’t like this theory. Although everyone else should be happier in these societies, the person in question who might or might not commit suicide should also be, on average, happier. There’s no reason to think that the average hedonic distance between potential suicides and their neighbors is higher in these areas. Indeed, given that Scandinavia – and many of the other happy societies – are also some of the most equal societies, I would expect an unusually low hedonic distance between people. And in fact, I notice that suicide rates by country are negatively correlated with inequality – that is, the more unequal the country, the lower the suicide rate (wow, I definitely don’t remember seeing that one in The Spirit Level.)

On the other hand, I can’t for the life of me think of a better theory, so whatever.

Other things that increase suicide rates, by the way, include springtime, nice weather, high levels of education, and very occasionally antidepressants. My father, a very hard-headed internist, makes fun of me for doing psychiatry because “the whole field is just common sense”, but sometimes it really isn’t.

So you should probably think very carefully before using a difference in depression or suicide rates to support your pet theory about which societies work better than others.

# These Are A Few (More) Of My (Least) Favorite Things

One year ago, I wrote Ten Things I Want To Stop Seeing On The Internet In 2014.

And now it’s 2015, and I think things are getting better. Take doge. I swear to God that the last time I saw the word doge, it was referring to an honest-to-God Venetian noble. And the price of dogecoin is down an order of magnitude from its peak last February. The War on Doges is starting to seem winnable.

I can’t take any credit for this. It has been a concerted effort on the part of millions of people who saw doge memes on Facebook, let their fingers briefly drift towards the “share” button, and then pulled themselves back from the precipice, restrained by their better nature.

But in the hopes that this is the first success of many, I would like to share some things I want to stop seeing on the Internet in 2015:

1. Abuse Of Poe’s Law

Poe’s Law is the belief that some religious fundamentalists are so stupid that it’s impossible to distinguish them from a parody.

This is all nice and well in the abstract, but when applied to a particular case, where a particular atheist has fallen for a parody site, it tends to be an unfortunate stand-in for “Some atheists are so ignorant that it’s impossible for them to distinguish religious people from a parody of religious people.” Listen:

A: “The Pope just said that everyone who isn’t creationist should be put in jail! What an outrage!”

B: “Uh, you do know that’s on The Onion, right?”

A: “Oh, well, haha, Poe’s Law, just goes to show how dumb those religious people are.”

Problem is, Poe’s Law isn’t limited to religion any more. Now it’s politics, culture, science, and anywhere else where one side thinks their opponents are so stupid it’s literally impossible to parody them (ie everywhere on both sides). You spread the dumbest and most obviously fake rumors to smear your opponents. And then when you’re caught, instead of admitting you were fooled, you claim Poe’s Law and smear your opponents even more.

On the other hand, once you’re willing to admit this dynamic exists, it can make for some pretty interesting guessing games and unintentional Intellectual Turing Tests – see the Poe’s Law In Action subreddit for some examples.

2. People Getting Destroyed By Other People

Whenever I write a persuasive piece, I get to see my fans share it on Twitter like this:



I didn’t destroy anybody. I disagreed with them.

I’m glad to know I’m not the only one who has to deal with this. Newsweek writes about how Jon Stewart Is A Violent Sociopath Who Must Be Stopped in reponse to increasing claims that Stewart “destroys”, “demolishes”, “disembowels”, and “makes ground beef” out of whoever he’s arguing against on his show.

This bothers me the same way that “debunked” bothers me. Both sides are going to insist that their own research “debunks” the other, and so make it impossible to have a conversation based on the premise that there’s still room for disagreement. The flip side of my fans believing that I’ve destroyed whoever is that when that person writes a response, their fans are going to believe they’ve destroyed me.

At least no one can eviscerate me, since Jon Stewart has already eviscerated the entire blogosphere.

3. Demonstrating That People Are Stupid By Having Them Use The Word “Muh”

No straw man is ever concerned about immigrants stealing his job. He’s always concerned about immigrants stealing “muh jarb”, or possibly “muh jawb”, which sounds like some form of obscure Islamic garment.

This has lately taken a disturbing turn in the form of straw feminists worrying about “muh sojiny”. I strongly believe that every women has a right to her sojiny and no man should be able to take it from her, but I still can’t help wishing that people would lay off the cheap shots for a while.

4. Wikipedia-Shaming

Did you know it is 2015 and people will still criticize you for getting facts off of Wikipedia?

I’m not even talking about controversial conclusions, like “on balance, the research about gun control shows…”. I’m talking about simple facts.

A: “China is bigger than the United States”

B: “Where’d you hear that one, Wikipedia?”

A: “…yes?”

B: “You expect me to believe something you literally just took off a Wikipedia article?”

Yes. Yes I do. I could go find the CIA World Factbook or whatever, but it will say the same thing as Wikipedia, because Wikipedia is pretty much always right. When you challenge Wikipedia on basic facts, all you do is force people to use inconvenient sources to back up the things Wikipedia says, costing people time for no reason and making them hate you. There may have been a time when Wikipedia was famously inaccurate. Or maybe there wasn’t. I don’t know. Wikipedia doesn’t have an article on it, so it would take time and energy to find out. The point is, now it’s 2015, and the matter has been settled.

How accurate is Wikipedia?:

Several studies have been done to assess the reliability of Wikipedia. An early study in the journal Nature said that in 2005, Wikipedia’s scientific articles came close to the level of accuracy in Encyclopædia Britannica and had a similar rate of “serious errors”. The study by Nature was disputed by Encyclopædia Britannica, and later Nature replied to this refutation with both a formal response and a point-by-point rebuttal of Britannica?’?s main objections. Between 2008 and 2012, articles in medical and scientific fields such as pathology, toxicology, oncology, pharmaceuticals, and psychiatry comparing Wikipedia to professional and peer-reviewed sources found that Wikipedia’s depth and coverage were of a high standard.

I know this because I got it from Wikipedia’s Reliability Of Wikipedia article. Go ahead, challenge me, I dare you.

5. Articles That Start Off With An Image Taking Up The Entire Screen

This is what I’m talking about. I click on the link expecting an article on gas pipeline deals, and there is exactly zero article on the first screenfull of page I come to. That’s fine, you know, the only reason I even clicked was to see a huge, high-resolution picture of Vladimir Putin’s head. Information is totally optional. Screw you. This is why if I want to learn about Russian-Chinese gas deals, I’ll just look them up on Wikipedia.

I feel the same way about those Web 2.0 sites where the landing page is just an image of a smiling group of people engaged in a nondescript activity, and then way up in the corner is a tiny button that says “Discover” (it’s always “discover”) which leads to actual information. Likewise this site, which probably made its designer feel very smug about their clean minimalist style, but where you can’t get a single word of information without watching a video.

6. Ads That Disappear Very Slowly

You get an ad. It appears at the bottom of the screen. You look at it, decide you’re not interested, click the little X. It disappears. But not right away. It crawls. It saunters. After what seems to be a long and arduous journey, during which it had to ford several rivers and stop off at Fort Laramie for supplies, it finally makes it to the bottom of the screen and fades away.

I try hard to understand other people’s perspectives. I know that companies need to have ads to make money. I know that they have an incentive to make those ads as disruptive and obnoxious as possible to make you look at them. I even understand why some ads have the little x kind of hidden, so you can’t find it without some poking around, which forces you to view the ad for a little while longer. I understand all those things.

But I don’t understand why the ad has to take so long to disappear. It’s obviously not just incompetence. They specifically have to add an extra little sliding-down animation to the ad to make it take so long. They put in more work to make it more annoying for no benefit. Do you really think that while I’m waiting for the ad to disappear, I’m thinking “You know, I thought I didn’t need to meet hot desperate singles in my area, which is why I clicked the X to make it go away, but that sliding-down-the-screen animation is so cool that I’m going to reload the page a couple of times, wait for the ad to come back, and then click it”?

7. Overuse Of Demonstratives In Clickbait

I understand that demonstratives (“this”, “that”, “yon”) are supposed to give you a bit of mystery, make you want to click on the article to see what’s happening. “This Celebrity Just Came Out As Gay” makes you wonder which one it is. “Rare Disease Spreads To These Three US States” makes you check if yours is one of them. Fine. I would personally prefer “Rare Disease Spreads To Three States” or even “Which Three States Did A Rare Disease Spread To? Click Here To Find Out”, but whatever.

However. On Vox recently, Obama Just Hit These North Koreans With Sanctions. What, exactly, are we supposed to get out of this? “Oh! I wonder if it’s Yu Kwang Ho! Surely they wouldn’t get Yu Kwang Ho! Better click to find out!”

8. Any Use Of The Word “Entitled”

Okay, I’ve already written on what I think of people calling nerds “entitled”. But it goes beyond that.

Number 8 in last year’s Least Favorite Things was “arguments about which generation is better”. Well, now those have progressed to arguments over which generation is most entitled. Hard Work? No Thanks! Meet Entitled-To-It-All Generation Y. Millenials are Selfish and Entitled and Helicopter Parents Are To Blame. But The Most Entitled Generation Isn’t Millennials, It’s Baby Boomers. And coming in from left field, maybe The Greatest Generation Was The Most Entitled. There are even entire books about this

Men feel entitled to women. Women feel entitled to men. Blacks feel entitled. Whites feel entitled. The Entitlement Mentality of Liberals coexists with Entitled Conservative White Male Putzes, possibly because Conservatives Feel ‘Entitled’ To Scorn ‘Entitlement’ (whatever).

Anyone can call their out-group entitled. The easiest way is – well, poor people are entitled because they demand hand-outs without working for them. Rich people are entitled because they think they deserve 100% of what they have and refuse to acknowledge or change the inequalities in the system that benefit them. One side or the other of that dichotomy is likely to map onto whatever group you want to insult.

“Entitled” is a Fully General Insult that can apply to anyone, and it really hurts. That makes it irresistable to the wrong kind of people, and it’s why I hope I start seeing less of it. Alternately, people could start giving their enemies the Psychological Entitlement Scale, which is so hilariously obvious with what it’s doing that I find it astounding that it apparently still manages to successfully detect some entitled people. The Titanic? Really?

9. People Being Post-Things

I recently heard someone describe themselves as “post-Zionist”, then go on to give what sounded like pretty standard criticism of Zionism. I don’t want to get too heavily into this particular example, because I understand post-Zionism is complex and every time I write something about Israel I get Israeli commenters saying I’ve gotten it wrong and other Israeli commenters saying no they’ve gotten it wrong and still other Israeli commenters saying we’ve all got it wrong. What was that saying about “two Jews, three opinions” again?

But what bothers me about post-Zionism is that it seems to carry this kind of smug “Oh, you guys are still Zionist? Don’t you know Zionism is, like, totally five years ago? Nowadays all the cool people have moved on to more exciting things,” which I don’t think really adds to the argument. Zionism versus anti-Zionism suggests a picture of two sides with two different opinions – which seems to match the reality pretty well. Zionism versus post-Zionism suggests one side just hasn’t gotten the message yet.

I feel the same way about post-rationalism. Yes, maybe you’ve seen through rationalism in some profound way and transcended it. Or maybe you just don’t get it. This is exactly the point under debate, and naming yourselves “post-rationalists” seems like an attempt to short-circuit it, not to mention leaving everyone else confused. And maybe you could give yourself a name that actually reflected your beliefs (“Kind Of New-Age-y People Who Are Better At Math Than Usual For That Demographic And Will Angrily Deny Being New-Age-y If Asked Directly”?) and we wouldn’t have to have a new “but what is post-rationalism?!?!” conversation every month.

Post-modernism can stay, though. At this point it’s less of a name than a warning label.

10. Disputes Over Whether Humans Evolved From Monkeys

I don’t mean creationism. I mean disputes among people who accept evolution, over whether it was monkeys in particular that humans evolved from.

It tends to go something like this.

A: “Humans evolved from monkeys”.

B: “No they didn’t! They evolved from chimps! Chimps are an ape, not a monkey!”

C: “Humans didn’t evolve from chimps! They evolved from a most recent common ancestor whose descendants include both humans and chimps!”

Everything about this conversation is not-even-wrong.

First, humans clearly evolved from monkeys in the same sense humans evolved from single-celled organisms. No one’s saying it had to be the most recent step.

Second, apes are ambiguously a type of monkey. Think square versus rectangle. All squares are rectangles but not all rectangles are squares and “rectangle” is usually used to indicate rectangles that are not squares but can technically refer to squares as well. Here’s a primatologist saying that Apes Are Monkeys – Deal With It.

Third, the most recent common ancestor of humans and chimpanzees may (or may not) have been a chimpanzee. This is Richard Wrangham’s thesis, and he calls it pan prior, placing it firmly within the chimpanzee genus.

These last two issues are especially annoying because they’re kind of meaningless category disputes. Yet for some reason the Internet seems to be obsessed with the lurking fear that someone, somewhere, might be saying that people evolved from monkeys or chimps.

Seriously. Get a life, Internet.

# Perceptions Of Required Ability Act As A Proxy For Actual Required Ability In Explaining The Gender Gap

I.

I briefly snarked about Leslie et al (2015) last week, but I should probably snark at it more rigorously and at greater length.

This is the paper that concludes that “women are underrepresented in fields whose practitioners believe that raw, innate talent is the main requirement for success because women are stereotyped as not possessing that talent.” They find that some survey questions intended to capture whether people believe a field requires innate talent correlate with percent women in that field at a fairly impressive level of r = -0.60.

The media, science blogosphere, et cetera has taken this result and run with it. A very small sample includes: National Science Foundation: Belief In Raw Brilliance May Decrease Diversity. Science Mag: the “misguided” belief that certain scientific fields require brilliance helps explain the underrepresentation of women in those fields. Reuters: Fields That Cherish Genius Shun Women. LearnU: Study Findings Point To Source Of Gender Gap In STEM. Scientific American: Hidden Hurdle Looms For Women In Science. Chronicle Of Higher Education: Disciplines That Expect Brilliance Tend To Punish Women. News Works: Academic Gender Gaps Tied To Stereotypes About Genius. Mathbabe: “The genius myth” keeps women out of science. Vocativ: Women Avoid Fields Full Of Self-Appointed Geniuses. And so on in that vein.

Okay. Imagine a study with the following methodology. You survey a bunch of people to get their perceptions of who is a smoker (“97% of his close friends agree Bob smokes”). Then you correlate those numbers with who gets lung cancer. Your statistics program lights up like a Christmas tree with a bunch of super-strong correlations. You conclude “Perception of being a smoker causes lung cancer”, and make up a theory about how negative stereotypes of smokers cause stress which depresses the immune system. The media reports that as “Smoking Doesn’t Cause Cancer, Stereotypes Do”.

This is the basic principle behind Leslie et al (2015).

The obvious counterargument is that people’s perceptions may be accurate, so your perception measure might be a proxy for a real thing. In the smoking study, we expect that people’s perception of smoking only correlates with lung cancer because it correlates with actual smoking which itself correlates with lung cancer. You would expect to find that perceived smoking correlates with lung cancer less than actual smoking, because the perceived smoking correlation is just the actual smoking correlation plus some noise resulting from misperceptions.

So I expected the paper to investigate whether or not perceived required ability correlated more, the same as, or less than actual required ability. Instead, they simply write:

Are women and African-Americans less likely to have the natural brilliance that some fields believe is required for top-level success? Although some have argued that this is so, our assessment of the literature is that the case has not been made that either group is less likely to possess innate intellectual talent1.

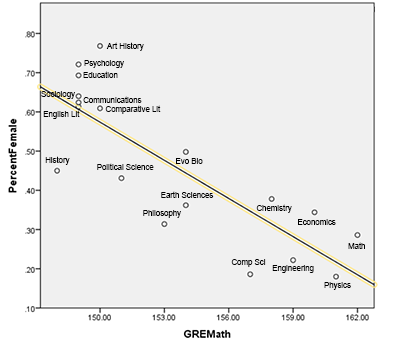
So we will have to do this ourselves. The researchers helpfully include in their supplement a list of the fields they studied and GRE scores for each, as part of some sub-analysis to check for selectivity. GRE scores correlate closely with IQ and with a bunch of measures of success in graduate school, so this sounds like it would be a good test of the actual required ability hypothesis. Let’s use this to figure out whether actual innate ability explains the discrepancies better or worse than perceived innate ability does.

When I use these data I find no effect of GRE scores on female representation.

But these data are surprising – for example, Computer Science had by far the lowest GRE score (and hence projected IQ?) of any field, which matches neither other sources nor my intuition. I looked more closely and found their measure combines Verbal, Quantitative, and Writing GREs. These are to some degree anti-correlated with each other across disciplines2; ie those disciplines whose students have higher Quantitative tend to have lower Writing scores (not surprising; consider a Physics department versus an English department).

Since the study’s analysis included two measures of verbal intelligence and only one measure of mathematical intelligence, it makes more mathematical departments appear to have lower scores and lower innate ability. Certainly a measure set up such that computer scientists get the lowest intelligence of everyone in the academy isn’t going to find innate ability related to STEM!

Since the gender gap tends to favor men in more mathematical subjects, if we’re checking for a basis in innate ability we should probably disentangle these tests and focus on the GRE Quantitative. I took GRE Quantitative numbers by department from the 2014 edition of the ETS report. The results looked like this:



There is a correlation of r = -0.82 (p = 0.0003) between average GRE Quantitative score and percent women in a discipline. This is among the strongest correlations I have ever seen in social science data. It is much larger than Leslie et al’s correlation with perceived innate ability3.

Despite its surprising size this is not a fluke. It’s very similar to what other people have found when attempting the same project. There’s a paper from 2002, Templer and Tomeo, that tries the same thing and finds r = 0.76, p < 0.001. Randal Olson tried a very similar project on his blog a while back and got r = 0.86. My finding is right in the middle.

A friendly statistician went beyond my pay grade and did a sequential ANOVA on these results4 and Leslie et al’s perceived-innate-ability results. They found that they could reject the hypothesis that the effect of actual innate ability was entirely mediated by perceived innate ability (p = 0.002), but could not reject the hypothesis that the effect of perceived-innate-ability was entirely mediated by actual-innate ability (p = 0.36).

In other words, we find no evidence for a continuing effect of people’s perceptions of innate ability after we adjust for what those perceptions say about actual innate ability, in much the same way we would expect to see no evidence for a continuing effect of people’s perceptions of smoking on lung cancer after we adjust for what those perceptions say about actual smoking.

II.

Correlation is not causation, but a potential causal mechanism can be sketched out.

I’m going to use terms like “ability” and “innate ability” and “genius” and “brilliance” because those are the terms Leslie et al use, but I should clarify. I’m using them the way Leslie et al seem to, as a contrast to hard work, the internal factors that give different people different payoffs per unit effort. So a genius is someone who can solve difficult problems with little effort; a dullard is one who can solve them only with great effort or not at all.

This use of “innate ability” is not the same thing as “genetically determined ability”. Genetically determined ability will be part of it, but there will also be many other factors. Environmental determinants of intelligence, like good nutrition and low lead levels. Exposure to intellectual stimulation during crucial developmental windows. The effect of steretoypes, insofar as those stereotypes globally decrease performance. Even previous training in a field might represent “innate ability” under this definition, although later we’ll try to close that loophole.

Academic programs presumably want people with high ability. The GRE bills itself as an ability test, and under our expanded definition of ability this is a reasonable claim. So let’s talk about what would happen if programs selected based solely on ability as measured by GREs.

This is, of course, not the whole story. Programs also use a lot of other things like grades, interviews, and publications. But these are all correlated with GRE scores, and anyway it’s nice to have a single number to work with. So for now let’s suppose colleges accept applicants based entirely on GRE scores and see what happens. The STEM subjects we’re looking at here are presumably most interested in GRE Quantitative, so once again we’ll focus on that.

Mathematics unsurprisingly has the highest required GRE Quantitative score. Suppose that the GRE score of the average Mathematics student – 162.0 – represents the average level that Mathematics departments are aiming for – ie you must be this smart to enter.

The average man gets 154.3 ± 8.6 on GRE Quantitative. The average woman gets 149.4 ± 8.1. So the threshold for Mathematics admission is 7.7 points ahead of the average male test-taker, or 0.9 male standard deviation units. This same threshold is 12.6 points ahead of the average female test-taker, or 1.55 female standard deviation units.

GRE scores are designed to follow a normal distribution, so we can plug all of this into our handy-dandy normal distribution calculator and find that 19% of men and 6% of women taking the GRE meet the score threshold to get into graduate level Mathematics. 191,394 men and 244,712 women took the GRE last year, so there will be about 36,400 men and 14,700 women who pass the score bar and qualify for graduate level mathematics. That means the pool of people who can do graduate Mathematics is 29% female. And when we look at the actual gender balance in graduate Mathematics, it’s also 29% female.

Vast rivers of ink have been spilled upon the question of why so few women are in graduate Mathematics programs. Are interviewers misogynist? Are graduate students denied work-life balance? Do stereotypes cause professors to “punish” women who don’t live up to their sexist expectations? Is there a culture of sexual harassment among mathematicians?

But if you assume that Mathematics departments are selecting applicants based on the thing they double-dog swear they are selecting applicants based on, there is literally nothing left to be explained5.

I am sort of cheating here. The exact perfect prediction in Mathematics is a coincidence. And I can’t extend this methodology rigorously to any other subject because I would need a much more complicated model where people of a given score level are taken out of the pool as they choose the highest-score-requiring discipline, leaving fewer high-score people available for the low-score-requiring ones. Without this more complicated task, at best I can set a maximum expected gender imbalance, then eyeball whether the observed deviation from that maximum is more or less than expected. Doing such eyeballing, there are slightly fewer women in graduate Physics and Computer Science than expected and slightly more women in graduate Economics than expected.

But on the whole, the prediction is very good. That it is not perfect means there is still some room to talk about differences in stereotypes and work-life balance and so on creating moderate deviations from the predicted ratio in a few areas like computer science. But this is arguing over the scraps of variance left over, after differences in mathematical ability have devoured their share.

III.

There are a couple of potentially very strong objections to this hypothesis. Let me see if I can answer them.

First, maybe this is a binary STEM vs. non-STEM thing. That is, STEM fields require more mathematical aptitude (obviously) and they sound like the sort to have more stereotypes about women. So is it possible that my supposedly large sample size is actually just showing an artifact of division into these two categories?

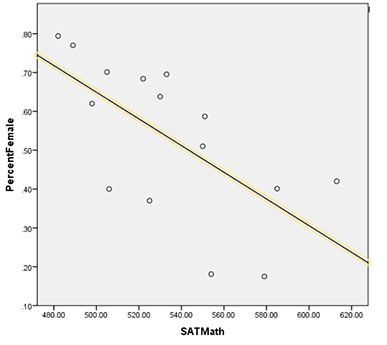
No. I divided the fields into STEM and non-STEM and ran an analysis within each subgroup. Within the non-STEM subgroup, there was a correlation between GRE Quantitative and percent female in a major of -0.64, p = 0.02. It is completely irresponsible to do this within the STEM subgroup, because it has n = 7 which is too small a sample size to get real results. But if we are bad people and do it anyway, we find a very similar correlation of -0.63. p is only 0.12, but with n=7 what did you expect?

Both of these correlations are higher than Leslie et al were able to get from their entire sample.

Second, suppose that it’s something else driving gender-based patterns in academia. Maybe stereotypes or long hours or whatever. Presumably, these could operate perfectly well in undergrad. So stereotypes cause lots of men to go into undergraduate math and lots of women to go into undergraduate humanities. The men in math classes successfully learn math and the women in humanities classes successfully learn humanities. Then at the end of their time in college they all take the GRE, and unsurprisingly the men who have been taking all the math classes do better in math. In this case, the high predictive power of mathematical ability would be a result of stereotypes, not an alternative to them.

In order to investigate this possibility we could look at SAT Math instead of GRE Quantitative scores, since these would show pre-college ability. SAT scores show a gap much like that in GRE scores; in both, the percentile of the average woman is in the low 40s.

Here is a graph of SAT Math scores against percent women in undergraduate majors:



SAT Math had a correlation of -0.65, p = 0.016.

This correlation is still very strong. It is still stronger than Leslie et al’s correlation with perceived required ability. But it is slightly weaker than the extremely strong correlation we find with GRE scores. Why?

I can’t answer that for sure, but here is a theory. The “undergraduate major” data is grabbed from what SAT test-takers put down as their preferred undergraduate major when they take the test in (usually) 11th grade. The “percent female” data is grabbed from records of degrees awarded in each field. So these are not exactly the same people on each side. One side shows the people who thought they wanted to do Physics in 11th grade. The other side shows the people who ended up completing a Physics degree.

The people who intend to pursue Physics but don’t end up getting a degree will be those who dropped out for some reason. While there are many reasons to drop out, one no doubt very common one is that the course was too hard. Therefore, the people who drop out will be disproportionately those with lower mathematical ability. Therefore, the average SAT Math score of 11th grade intended Physics majors will be lower than the average SAT Math score of Physics degree earners. So the analysis above likely underestimates the average SAT Math score of people in mathematical fields. This could certainly explain the lower correlation, and I predict that if we could replace our unrepresentative measure of SAT scores with a more representative one, much of the gap between this correlation and the previous one would close.

These data do not rule out simply pushing everything back a level and saying that these stereotypes affect what classes girls take in middle school and high school. Remember, we using “ability” as a designation for a type of excellence, not an explanatory theory of it. This simply confirms that by eleventh grade, the gap has already formed.7.

Third, perhaps SAT and GRE math tests are not reflective of women’s true mathematical ability. This is the argument from stereotype threat, frequently brought up as reasons why tests should not be used to judge aptitude.

But this is based on a fundamental misunderstanding of stereotype threat found in the popular media, which actual researchers in the field keep trying to correct (to no avail). See for example Sackett, Hardison, and Cullen (2004), who point out that no research has ever claimed stereotype threat accounts for gender gaps on mathematics tests. What the research found was that, by adding an extra stereotype threat condition, you could widen those gaps further. The existing gaps on tests like the SAT and GRE correspond to the “no stereotype threat” control condition in stereotype threat experiments, and “absent stereotype threat, the two groups differ to the degree that would be expected based on differences in prior SAT scores”. Aronson and Steele, who did the original stereotype threat research and invented the field, have confirmed that this is accurate and endorsed the warning.

Anyway, even if the pop sci version of stereotype threat were entirely true and explained everything, it still wouldn’t rescue claims of bias or sexism in the sciences. It would merely mean that the sciences’ reasonable and completely non-sexism-motivated policy of trusting test scores was ill-advised.8

Fourth, might there be reverse causation? That is, suppose that there are stereotypes and sexism restricting women’s entry into STEM fields, and unrelatedly men have higher test scores. Then the fields with the stereotypes would end up with the people with higher test scores, and it would look like they require more ability. Might that be all that’s happening here?

No. I used gender differences in the GRE scores to predict what scores we would expect each major to have if score differences came solely from differences in gender balance. This predicted less than a fifth of the variation. For example, the GRE Quantitative score difference between the average test-taker and the average Physics graduate student was 9 points, but if this were solely because of differential gender balance plus the male test advantage we would predict a difference of only 1.5 points. The effect on SAT scores is similarly underwhelming.

But I think the most important thing I want to say about objections to Part II is that, whether they’re correct or not, Part I still stands. Even if the correlation between innate ability and gender balance turns out to be an artifact, Leslie et al’s correlation between perceived innate ability and gender balance is still an artifact of an artifact.

IV.

A reader of an early draft of this post pointed out the imposingly-named Nonlinear Psychometric Thresholds In Physics And Mathematics. This paper uses SAT Math scores and GPA to create a model in which innate ability and hard work combine to predict the probability that a student will be successful in a certain discipline. It finds that in disciplines “such as Sociology, History, English, and Biology” these are fungible – greater work ethic can compensate for lesser innate ability and vice versa. But in disciplines such as Physics and Mathematics, this doesn’t happen. People below a certain threshold mathematical ability will be very unlikely to succeed in undergraduate Physics and Mathematics coursework no matter how hard-working they are.

And that brought into relief part of why this study bothers me. It ignores the pre-existing literature on the importance of innate ability versus hard work. It ignores the rigorous mathematical techniques developed to separate innate ability from hard work. Not only that, but it ignores pre-existing literature on predicting gender balance in different fields, and the pre-existing literature on GRE results and what they mean and how to use them, and all the techniques developed by people in those areas.

Having committed itself to flying blind, it takes the thing we already know how use to predict gender balance, shoves it aside in favor of a weird proxy for that thing, and finds a result mediated by that thing being a proxy for the thing they are inexplicably ignoring. Even though it just used a proxy for aptitude to predict gender balance, everyone congratulates it for having proven that aptitude does not affect gender balance.

Science journalism declares that the myth that ability matters has been vanquished forever. The media take the opportunity to remind us that scientists are sexist self-appointed geniuses who use stereotypes to punish women. And our view of an important issue becomes just a little muddier.

I encourage everyone to reanalyze this data and see if I’m missing something. You can find the GRE data I used here and the SAT data here (both in .xlsx format).

Footnotes

1. They cite for this claim, among other things, Stephen Jay Gould’s The Mismeasure Of Man

2. Beware the ecological fallacy; these scores are still positively correlated in individuals.

3. It was also probably more highly significant, but I can’t tell for sure because (ironically) their significance result wasn’t to enough significant digits.

4. There was a small error in the percent of women in Communications in the dataset I provided them with, so these numbers are off by a tiny fraction from what you will get if you try to replicate. I didn’t feel comfortable asking them to redo the entire thing, but the small error would not have changed the results significantly, and the tiny amount it would have changed them would have been in the direction of making the innate ability results more striking rather than less.

5. Although Leslie et al focused on women, they believe their results could also extend to why African-Americans are underrepresented compared to European-Americans and Asian-Americans in certain subjects. They theorize that European and Asian Americans, like men, are stereotyped as innately brilliant, but African-Americans, like women, lack this stereotype. I find this a bit off – after all, in the gender results, they contrasted the male “more innately brilliant” stereotype with the female “harder-working” stereotype, but African Americans suffer from a stereotype of not being hard-working, and Asian-Americans do have a stereotype of being hard-working, even more so than women. Anyway, this is only a mystery if you stick to Leslie et al’s theory of stereotypes about perceived innate ability. Once you look at GRE Quantitative scores, you find that whites average 150.8, Asians average 153.9, and blacks average 143.7, and there’s not much left to explain.

6. It’s hard to correlate SAT scores with majors, because the SAT data is full of tiny vocational majors that throw off the results. For example, there are two hundred people in the country studying some form of manufacturing called “precision production”, they’re almost all male, and they have very low SAT scores. On the other hand, there are a few thousand people studying something called “family science”, they’re almost all women, and they also all have very low SAT scores. The shape of gender\*major\*SAT scores depends almost entirely on how many of these you count. I circumvented the entire problem by just counting the fields that approximately corresponded to the ones Leslie et al counted in their graduate-level study. I tried a few different analyses using different ways of deciding which fields to count, and as long as they were vaguely motivated by a desire to include academic subjects and not the vocational subjects with very low scores, they all came out about the same.

7. The argument that stereotypes cause boys to take more middle school and high school math classes than girls is somewhat argued against by the finding that actually girls take more middle school and high school math classes than boys. However, there are some contrary results; for example, boys are more likely than girls to take the AP Calculus test. This entire area gets so tangled up in differing levels of interest and ability and work-ethic that it’s not worth it, at my level of interest and ability and work ethic, to try to work it out. The best I can say is that the gap appears by the time kids take the SAT in 11th grade.

8. I can’t help adding that I continue to believe that the stereotype threat literature looks like a null field which continues to exist only through publication bias and experimenter effects. The funnel plot shows a clear peak at “zero effect” and an asymmetry indicating a publication bias for positive results (for some discussion of why I like funnel plots, see here.) And a closer look at the individual research shows this really disturbing pattern of experiments by true believers finding positive effects, experiments by neutral parties and skeptics not finding them, replication attempts failing, and large real-world quasi-experiments turning up nothing – in a way very reminiscent of parapsychology. Although I am far from 100% sure, I would tentatively place my money on the entire idea of stereotype threat vanishing into the swamp of social psychology’s crisis of replication.

# A Philosopher Walks Into A Coffee Shop

I have been really enjoying literarystarbucks.tumblr.com, which publishes complicated jokes about what famous authors and fictional characters order at Starbucks. I like it so much I wish I knew more great literature, so I could get more of the jokes.

Since the creators seem to be restricting themselves to the literary world, I hope they won’t mind if I fail to resist the temptation to steal their technique for my own field of interest. Disclaimer: two of these are widely-known philosophy jokes and not original to me.

\* \* \*

Parmenides goes up to the counter. “Same as always?” asks the barista. Parmenides nods.

\* \* \*

Pythagoras goes up to the counter and orders a caffe Americano. “Mmmmm,” he says, tasting it. “How do you guys make such good coffee?” “It’s made from the freshest beans,” the barista answers. Pythagoras screams and runs out of the store.

\* \* \*

Thales goes up to the counter, says he’s trying to break his caffeine habit, and orders a decaf. The barista hands it to him. He takes a sip and spits it out. “Yuck!” he says. “What is this, water?”

\* \* \*

Gottfried Leibniz goes up to the counter and orders a muffin. The barista says he’s lucky since there is only one muffin left. Isaac Newton shoves his way up to the counter, saying Leibniz cut in line and he was first. Leibniz insists that he was first. The two of them come to blows.

\* \* \*

Georg Wilhelm Friedrich Hegel goes up to the counter and gives a tremendously long custom order in German, specifying exactly how much of each sort of syrup he wants, various espresso shots, cream in exactly the right pattern, and a bunch of toppings, all added in a specific order at a specific temperature. The barista can’t follow him, so just gives up and hands him a small plain coffee. He walks away. The people behind him in line are very impressed with his apparent expertise, and they all order the same thing Hegel got. The barista gives each of them a small plain coffee, and they all remark on how delicious it tastes and what a remarkable coffee connoisseur that Hegel is. “The Hegel” becomes a new Starbucks special and is wildly popular for the next seventy years.

\* \* \*

Socrates goes up to the counter. “What would you like?” asks the barista. “What would you recommend?” asks Socrates. “I would go with the pumpkin spice latte,” says the barista. “Why?” asks Socrates. “It’s seasonal,” she answers. “But why exactly is a seasonal drink better than a non-seasonal drink?” “Well,” said the barista, “I guess it helps to connect you to the rhythm of the changing seasons.” “But do you do other things to connect yourself to that rhythm?” asked Socrates. “Like wear seasonal clothing? Or read seasonal books? If not, how come it’s only drinks that are seasonal?” “I’m not sure,” says the barista. “Think about it,” says Socrates, and leaves without getting anything.

\* \* \*

Rene Descartes goes up to the counter. “I’ll have a scone,” he says. “Would you like juice with that?” asks the barista. “I think not,” says Descartes, and he ceases to exist.

\* \* \*

Jean-Paul Sartre goes up to the counter. “What do you want?” asks the barista. Sartre thinks for a long while. “What do? I want?” he asks, and wanders off with a dazed look on his face.

\* \* \*

William of Occam goes up to the counter. He orders a coffee.

\* \* \*

Adam Smith goes up to the counter. “I’ll have a muffin,” he says. “Sorry,” says the barista, “but those two are fighting over the last muffin.” She points to Leibniz and Newton, who are still beating each other up. “I’ll pay $2 more than the sticker price, and you can keep the extra,” says Smith. The barista hands him the muffin.

\* \* \*

John Buridan goes up to the counter and stares at the menu indecisively.

\* \* \*

Ludwig Wittgenstein goes up to the counter. “I’ll have a small toffee mocha,” he says. “We don’t have small,” says the barista. “Then what sizes do you have?” “Just tall, grande, and venti.” “Then doesn’t that make ‘tall’ a ‘small’?” “We call it tall,” says the barista. Wittgenstein pounds his fist on the counter. “Tall has no meaning separate from the way it is used! You are just playing meaningless language games!” He storms out in a huff.

\* \* \*

St. Anselm goes up to the counter and considers the greatest coffee of which it is possible to conceive. Since existence is more perfect than nonexistence, the coffee must exist. He brings it back to his table and drinks it.

\* \* \*

Ayn Rand goes up to the counter. “What do you want?” asks the barista. “Exactly the relevant question. As a rational human being, it is my desires that are paramount. Since as a reasoning animal I have the power to choose, and since I am not bound by any demand to subordinate my desires to that of an outside party who wishes to use force or guilt to make me sacrifice my values to their values or to the values of some purely hypothetical collective, it is what I want that is imperative in this transaction. However, since I am dealing with you, and you are also a rational human being, under capitalism we have an opportunity to mutually satisfy our values in a way that leaves both of us richer and more fully human. You participate in the project of affirming my values by providing me with the coffee I want, and by paying you I am not only incentivizing you for the transaction, but giving you a chance to excel as a human being in the field of producing coffee. You do not produce the coffee because I am demanding it, or because I will use force against you if you do not, but because it most thoroughly represents your own values, particularly the value of creation. You would not make this coffee for me if it did not serve you in some way, and therefore by satisfying my desires you also reaffirm yourself. Insofar as you make inferior coffee, I will reject it and you will go bankrupt, but insofar as your coffee is truly excellent, a reflection of the excellence in your own soul and your achievement as a rationalist being, it will attract more people to your store, you will gain wealth, and you will be able to use that wealth further in pursuit of excellence as you, rather than some bureaucracy or collective, understand it. That is what it truly means to be a superior human.” “Okay, but what do you want?” asks the barista. “Really I just wanted to give that speech,” Rand says, and leaves.

\* \* \*

Voltaire goes up to the counter and orders an espresso. He takes it and goes to his seat. The barista politely reminds him he has not yet paid. Voltaire stays seated, saying “I believe in freedom of espresso.”

\* \* \*

Thomas Malthus goes up to the counter and orders a muffin. The barista tells him somebody just took the last one. Malthus grumbles that the Starbucks is getting too crowded and there’s never enough food for everybody.

\* \* \*

Immanuel Kant goes up to the counter at exactly 8:14 AM. The barista has just finished making his iced cinnamon dolce latte, and hands it to him. He sips it for eight minutes and thirty seconds, then walks out the door.

\* \* \*

Bertrand Russell goes up to the counter and orders the Hegel. He takes one sip, then exclaims “This just tastes like plain coffee! Why is everyone making such a big deal over it?”

\* \* \*

Pierre Proudhon goes up to the counter and orders a Tazo Green Tea with toffee nut syrup, two espresso shots, and pumpkin spice mixed in. The barista warns him that this will taste terrible. “Pfah!” scoffs Proudhon. “Proper tea is theft!”

\* \* \*

Sigmund Freud goes up to the counter. “I’ll have ass sex, presto,” he says. “What?!” asks the barista. “I said I’ll have iced espresso.” “Oh,” said the barista. “For a moment I misheard you.” “Yeah,” Freud tells her. “I fucked my mother. People say that.” “WHAT?!” asks the barista. “I said, all of the time other people say that.”

\* \* \*

Jeremy Bentham goes up to the counter, holding a $50 bill. “What’s the cheapest drink you have?” he asks. “That would be our decaf roast, for only $1.99,” says the barista. “Good,” says Bentham and hands her the $50. “I’ll buy those for the next twenty-five people who show up.”

\* \* \*

Patricia Churchland walks up to the counter and orders a latte. She sits down at a table and sips it. “Are you enjoying your beverage?” the barista asks. “No,” says Churchland.

\* \* \*

Friedrich Nietzsche goes up to the counter. “I’ll have a scone,” he says. “Would you like juice with that?” asks the barista. “No, I hate juice,” says Nietzsche. The barista misinterprets him as saying “I hate Jews”, so she kills all the Jews in Europe.

# The Parable Of The Talents

[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%s 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 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 talenton, 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]

# Talents Part 2: Attitude vs. Altitude

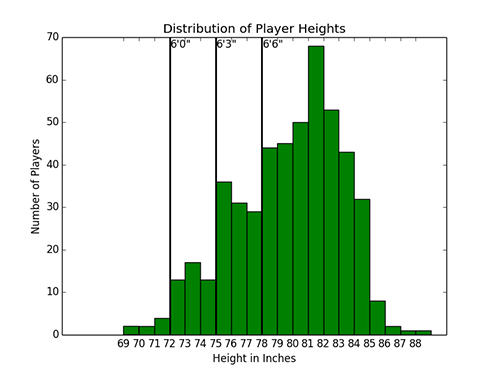
[Content note: scrupulosity and self-esteem triggers, IQ. Potentially not good for growth mindset.]

I.

The average eminent theoretical physicist has an IQ of 150-160. The average NBA player has a height of 6’7. Both of these are a little over three standard deviations above their respective mean. Since z-scores are magic and let us compare unlike domains, we conclude that eminent theoretical physicists are about as smart as pro basketball players are tall.

Any time people talk about intelligence, height is a natural sanity check. It’s another strongly heritable polygenic trait which is nevertheless susceptible to environmental influences, and which varies in a normal distribution across the population – but which has yet to accrete the same kind of cloud of confusion around it that IQ has.

So let’s see what we can learn from the height distribution of every player in the NBA: (source)



The first thing we notice is that nobody shorter than average (ie 5’9) gets in. This is not true eternally and forever – Wikipedia’s List Of Shortest NBA Players names 25 NBA players in history (out of maybe 5000 total) who were below average stature. The all-time record holder is Tyrone “Muggsy” Bogues at 5’3, who was actually pretty good (though not above exploiting his distinction for publicity: his autobiography is called In the Land of Giants). But Bogues was a spectacular outlier. If we’re going to stick to our resolution to use the histogram we’ve got, the contribution of the left half of the height bell curve is precisely zero.

We can do some very rough Fermi calculations for the next couple standard deviations.

34% of the US male population is 5’9 to 6’0, so about 54 million men. There are 5 NBA players in this band. So about one in every 11 million people of this height is in the NBA.

13.5% of the US male population is 6’0 to 6’3, so about 21 million men. There are 40 NBA players in this band. So about one in every 500,000 people of this height is in the NBA.

2.35% of the US male population is 6’3 to 6’6, so about 4 million men. There are 95 NBA players in this band. So about one in every 40,000 people of this height is in the NBA.

0.15% of the US male population is 6’6 to 6’9, so about 200,000 men. There are 130 NBA players in this band. So about one in every 1,500 people of this height is in the NBA.

0.003% of the US male population is 6’9 to 7’0, so about 5,000 men. There are 160 NBA players in this band. So about one in every 30 people of this height is in the NBA.

0.00002% of the US male population is 7’0 to 7’3, which corresponds to about 45 men. There are 40 NBA players in this band. So about 8 out of 9 people of this height are in the…wait, no, that can’t be right.

Sports Illustrated‘s Dan Diamond does a similar analysis and gets broadly similar results. But he adds several complicating factors. First, at this height people with very rare medical conditions and pituitary tumors start taking over from normal variation, so we lose our ability to derive a census a priori with pure math. Second of all, at this height talent scouts comb the world for suitably tall foreigners and import them, so we can no longer assume we’re drawing from the pool of tall Americans. And third of all, the same way you round up so that you’re 6’0 on OKCupid, NBA players round up so that they’re 7’0 on their official stats.

He concludes that most likely about 17% of seven-foot-tall young men in America are in the NBA. This might still be an overestimate, but is probably in the right ballpark. Forbes Magazine writes that Being Seven Feet Tall Is The Fastest Way To Get Rich In America and quotes a talent scout who says “that he’ll ‘check up on anyone over 7 feet that’s breathing.'” Given that a lot of people this height are unhealthy or uninterested, it might not be much of an exaggeration to say that if you’re 7’3 and have any interest in basketball, you’ve got better than even odds of going pro.

But why stop there? 0.000000001% of men are 7’3 to 7’6. Given the size of the American male population, there shouldn’t be a single person in the US with this height, though in practice a few people with endocrine abnormalities make the cut. There will, however, be a couple of healthy people of this height in the world population. There is one person currently in the NBA above 7’3, and he is a Tanzanian native discovered by talent scouts.

WE’VE GOT TO GO DEEPER TALLER! 0.00000000000001% of the population is 7’6 to 7’9. Statistically, there should not be a single man this tall in the entire world. No NBA players are currently this tall. But Yao Ming, who retired four years ago, was 7’6 exactly. He was the product of a Maoist breeding program specifically aimed at producing tall people to play basketball. You can break a lot of statistical laws if you have breeding programs and flexible ethics. Also if you have pituitary tumors, as the remainder of the List Of Tallest People reminds us. It looks like a little over half of the living people in this height band have played professional basketball at some point.

There are people taller than 7’9, but a lot of them have trouble standing without support, which somewhat decreases basketball ability.



On the plus side, you can do all sorts of awesome things, like take a picture standing next to the world’s shortest man, or…okay, pretty much just that.

II.

This offers us an opportunity to compare our intuitions about intellect to our intuitions about basketball. I recommend taking that opportunity. Basketball isn’t mysterious, it isn’t politicized, and it isn’t tied up with our notions of self-worth. As a result, our intuitions about basketball are crystal clear.

If somebody said “Height plays no role in your success as a basketball player,” we’d look at them funny. Literally nobody shorter than average is currently in the NBA. Your odds at average height are only one in ten million. But if you’re 7’0, you can pretty much walk right in.

But if somebody said “Well, success in basketball is clearly dependent on height, that means there’s no point in practicing, after all you’re not going to grow any taller,” we’d look at them funny too. I don’t know much about sports, but I predict a short guy who’s been playing for years can defeat an untrained giant every time. And of course if two giants go up against each other, experience will be decisive.

If someone found a 5’8 kid who really liked playing hoops in his backyard and was pretty good at it, and they told him “You’re a moron, you can never succeed at basketball because of your height, give up,” that person would be a jerk. The kid is enjoying himself, there’s no point in insulting him about it, and if his goal is just to end up as a college athlete or a minor leaguer, then with hard work he could certainly make it. Even if he didn’t, he might be able to apply what he’d learned to a related area like coaching or reporting.

But if that same kid wanted to go deeply into debt to attend a basketball training camp, and he’ll count it a failure if he achieves anything less than NBA superstardom, you should probably warn him that his hopes aren’t very realistic and that maybe he should lower his standards or pick a sport more suited to his body type or try another line of work entirely.

If somebody who was 6’6 complained that they’d never be able to beat the 7’0 players on the other side, we would tell them to brighten up. After all, Michael Jordan was “only” 6’6, and he’s maybe the greatest basketball player of all time, even though he often had to face off against people taller than himself.

But if a team made entirely of 6’6 players faced off against a team entirely of 7’0 players, and both of them were really motivated and practiced really hard and had great coaches, I know who I would bet on.

Most important, people all over the world innocently enjoy playing basketball, and they are right to do so. Everyone knows that taller players have an advantage, no one’s denying it. But at the levels most people play at, moderate height differences are surmountable by differences in training and technique, and large height differences are so rare as to not come up very often. Yes, the pro leagues are a different story. Yet somehow the 5’8 kid who scores a three-pointer still manages to feel good about himself. And we can be very impressed by Kobe Bryant and LeBron James, praise them for their determination and technique and competitive spirit, all while acknowledging that it’s not a coincidence that the one is 6’6 and the other 6’9 and neither of these is a normal-person height.

All of this is crystal clear when we’re talking about basketball. But as soon as we switch back to talking about intelligence, we’re shouting at each other: “YOU SAY RAMANUJAN WAS REALLY HARD-WORKING, BUT I THINK HE HAD HIGH IQ, SO THERE!” Or “WELL IF SUCCESS COMES FROM INNATE TALENT, I GUESS I’LL JUST NEVER STUDY AGAIN.”

Luke Muehlhauser liked to call his philosophy of religion “common sense atheism”, meaning that he wanted to treat the question of God with the same “common” reasoning that he used for every other question. If we don’t see a tiger in front yard, we don’t say “Since it’s impossible to prove a negative, I can at best be agnostic about the existence of a tiger”, we say “I guess there’s probably not a tiger.”

Likewise, if we can just apply the same common reasoning we use for normal everyday activities like basketball to the question of intelligence, we might find it’s not so complicated and scary after all.

Which is not to say that it’s not important. After all, the other analogy between intelligence and basketball talent is that they’re both skills we need to cultivate at the highest levels if we want to save the world when it is threatened by dangerous future technology we can barely comprehend.

[still more on this subject later, but not immediately]

# Practically-A-Book Review: Dying To Be Free

I am the last person with a right to complain about Internet articles being too long. But if I did have that right, I think I would exercise it on Dying To Be Free, the Huffington Post’s 20,000-word article on the current state of heroin addiction treatment. I feel like it could have been about a quarter the size without losing much.

It’s too bad that most people will probably shy away from reading it, because it gets a lot of stuff really right.

The article’s thesis is also its subtitle: “There’s a treatment for heroin addiction that actually works; why aren’t we using it?” To save you the obligatory introductory human interest story: that treatment is suboxone. Its active ingredient is the drug buprenorphine, which is kind of like a safer version of methadone. Suboxone is slow-acting, gentle, doesn’t really get people high, and is pretty safe as long as you don’t go mixing it with weird stuff. People on suboxone don’t experience opiate withdrawal and have greatly decreased cravings for heroin. I work at a hospital that’s an area leader in suboxone prescription, I’ve gotten to see it in action, and it’s literally a life-saver.

Conventional heroin treatment is abysmal. Rehab centers aren’t licensed or regulated and most have little interest in being evidence-based. Many are associated with churches or weird quasi-religious groups like Alcoholics Anonymous. They don’t necessarily have doctors or psychologists, and some actively mistrust them. All of this I knew. What I didn’t know until reading the article was that – well, it’s not just that some of them try to brainwash addicts. It’s more that some of them try to cargo cult brainwashing, do the sorts of things that sound like brainwashing to them, without really knowing how brainwashing works assuming it’s even a coherent goal to aspire to. Their concept of brainwashing is mostly just creating a really unpleasant environment, yelling at people a lot, enforcing intentionally over-strict rules, and in some cases even having struggle-session-type-things where everyone in the group sits in a circle, scream at the other patients, and tell them they’re terrible and disgusting. There’s a strong culture of accusing anyone who questions or balks at any of it of just being an addict, or “not really wanting to quit”.

I have no problem with “tough love” when it works, but in this case it doesn’t. Rehab programs make every effort to obfuscate their effectiveness statistics – I blogged about this before in Part II here – but the best guesses by outside observers is that for a lot of them about 80% to 90% of their graduates relapse within a couple of years. Even this paints too rosy a picture, because it excludes the people who gave up halfway through.

Suboxone treatment isn’t perfect, and relapse is still a big problem, but it’s a heck of a lot better than most rehabs. Suboxone gives people their dose of opiate and mostly removes the biological half of addiction. There’s still the psychological half of addiction – whatever it was that made people want to get high in the first place – but people have a much easier time dealing with that after the biological imperative to get a new dose is gone. Almost all clinical trials have found treatment with methadone or suboxone to be more effective than traditional rehab. Even Cochrane Review, which is notorious for never giving a straight answer to anything besides “more evidence is needed”, agrees that methadone and suboxone are effective treatments.

Some people stay on suboxone forever and do just fine – it has few side effects and doesn’t interfere with functioning. Other people stay on it until they reach a point in their lives when they feel ready to come off, then taper down slowly under medical supervision, often with good success. It’s a good medication, and the growing suspicion it might help treat depression is just icing on the cake.

There are two big roadblocks to wider use of suboxone, and both are enraging.

The first roadblock is the #@$%ing government. They are worried that suboxone, being an opiate, might be addictive, and so doctors might turn into drug pushers. So suboxone is possibly the most highly regulated drug in the United States. If I want to give out OxyContin like candy, I have no limits but the number of pages on my prescription pad. If I want to prescribe you Walter-White-level quantities of methamphetamine for weight loss, nothing is stopping me but common sense. But if I want to give even a single suboxone prescription to a single patient, I have to take a special course on suboxone prescribing, and even then I am limited to only being able to give it to thirty patients a year (eventually rising to one hundred patients when I get more experience with it). The (generally safe) treatment for addiction is more highly regulated than the (very dangerous) addictive drugs it is supposed to replace. Only 3% of doctors bother to jump through all the regulatory hoops, and their hundred-patient limits get saturated almost immediately. As per the laws of suppy and demand, this makes suboxone prescriptions very expensive, and guess what social class most heroin addicts come from? Also, heroin addicts often don’t have access to good transportation, which means that if the nearest suboxone provider is thirty miles from their house they’re out of luck. The List Of Reasons To End The Patient Limits On Buprenorphine expands upon and clarifies some of these points.

(in case you think maybe the government just honestly believes the drug is dangerous – nope. You’re allowed to prescribe without restriction for any reason except opiate addiction)

The second roadblock is the @#$%ing rehab industry. They hear that suboxone is an opiate, and their religious or quasi-religious fanaticism goes into high gear. “What these people need is Jesus and/or their Nondenominational Higher Power, not more drugs! You’re just pushing a new addiction on them! Once an addict, always an addict until they complete their spiritual struggle and come clean!” And so a lot of programs bar suboxone users from participating.

This doesn’t sound so bad given the quality of a lot of the programs. Problem is, a lot of these are closely integrated with the social services and legal system. So suppose somebody’s doing well on suboxone treatment, and gets in trouble for a drug offense. Could be that they relapsed on heroin one time, could be that they’re using something entirely different like cocaine. Judge says go to a treatment program or go to jail. Treatment program says they can’t use suboxone. So maybe they go in to deal with their cocaine problem, and by the time they come out they have a cocaine problem and a heroin problem.

And…okay, time for a personal story. One of my patients is a homeless man who used to have a heroin problem. He was put on suboxone and it went pretty well. He came back with an alcohol problem, and we wanted to deal with that and his homelessness at the same time. There are these organizations called three-quarters houses – think “halfway houses” after inflation – that take people with drug problems and give them an insurance-sponsored place to live. But the catch is you can’t be using drugs. And they consider suboxone to be a drug. So of about half a dozen three-quarters houses in the local area, none of them would accept this guy. I called up the one he wanted to go to, said that he really needed a place to stay, said that without this care he was in danger of relapsing into his alcoholism, begged them to accept. They said no drugs. I said I was a doctor, and he had my permission to be on suboxone. They said no drugs. I said that seriously, they were telling me that my DRUG ADDICTED patient who was ADDICTED TO DRUGS couldn’t go to their DRUG ADDICTION center because he was on a medication for treating DRUG ADDICTION? They said that was correct. I hung up in disgust.

So I agree with the pessimistic picture painted by the article. I think we’re ignoring our best treatment option for heroin addiction and I don’t see much sign that this is going to change in the future.

But the health care system not being very good at using medications effectively isn’t news. I also thought this article was interesting because it touches on some of the issues we discuss here a lot:

The value of ritual and community. A lot of the most intelligent conservatives I know base their conservativism on the idea that we can only get good outcomes in “tight communities” that are allowed to violate modern liberal social atomization to build stronger bonds. The Army, which essentially hazes people with boot camp, ritualizes every aspect of their life, then demands strict obedience and ideological conformity, is a good example. I do sometimes have a lot of respect for this position. But modern rehab programs seem like a really damning counterexample. If you read the article, you will see that this rehabs are trying their best to create a tightly-integrated religiously-inspired community of exactly that sort, and they have abilities to control their members and force their conformity – sometimes in ways that approach outright abuse – that most institutions can’t even dream of. But their effectiveness is abysmal. The entire thing is for nothing. I’m not sure whether this represents a basic failure in the idea of tight communities, or whether it just means that you can’t force them to exist ex nihilo over a couple of months. But I find it interesting.

My love-hate relationship with libertarianism. Also about the rehabs. They’re minimally regulated. There’s no credentialing process or anything. There are many different kinds, each privately led, and low entry costs to creating a new one. They can be very profitable – pretty much any rehab will cost thousands of dollars, and the big-name ones cost much more. This should be a perfect setup for a hundred different models blooming, experimenting, and then selecting for excellence as consumers drift towards the most effective centers. Instead, we get rampant abuse, charlatanry, and uselessness.

On the other hand, when the government rode in on a white horse to try to fix things, all they did was take the one effective treatment, regulate it practically out of existence, then ride right back out again. So I would be ashamed to be taking either the market’s or the state’s side here. At this point I think our best option is to ask the paraconsistent logic people to figure out something that’s neither government nor not-government, then put that in charge of everything.

Society is fixed, biology is mutable. People have tried everything to fix drug abuse. Being harsh and sending drug users to jail. Being nice and sending them to nice treatment centers that focus on rehabilitation. Old timey religion where fire-and-brimstone preachers talk about how Jesus wants them to stay off drugs. Flaky New Age religion where counselors tell you about how drug abuse is keeping you from your true self. Government programs. University programs. Private programs. Giving people money. Fining people money. Being unusually nice. Being unusually mean. More social support. Less social support. This school of therapy. That school of therapy. What works is just giving people a chemical to saturate the brain receptor directly. We know it works. The studies show it works. And we’re still collectively beating our heads against the wall of finding a social solution.

# Everything Not Obligatory Is Forbidden

[seen on the New York Times’ editorial page, February 6 2065, written by one “Dr. Mora LeQuivalence”]

It’s 2065. Not giving your kids super-enhancement designer baby gene therapy isn’t your “choice”. If you don’t super-enhance your kids, you are a bad parent. It’s that simple.

Harsh? Maybe. But consider the latest survey, which found that about five percent of parents fail to super-enhance their children by the time they enter kindergarten. These aren’t poor people who can’t afford super-enhancement designer baby gene therapy. These are mostly rich, highly educated individuals in places like California and Oregon who say they think it’s more “natural” to leave their children defenseless against various undesirable traits. “I just don’t think it’s right to inject retroviral vectors into my baby’s body to change her from the way God made her,” one Portland woman was quoted by the Times as saying earlier this week. Other parents referred to a 2048 study saying the retroviral injections, usually given in the first year of life, increase the risk of various childhood cancers – a study that has since been soundly discredited.

These parents will inevitably bring up notions of “personal freedom”. But even if we accept the dubious premise that parents have a right to sacrifice their children’s health, refusing super-enhancement designer baby gene therapy isn’t just a personal choice. It’s a public health issue that affects everybody in society.

In 2064 there were almost 200 murders nationwide, up from a low of fewer than 50 in 2060. Why is this killer, long believed to be almost eradicated, making a comeback? Criminologists are unanimous in laying the blame on unenhanced children, who lack the improved impulse-control and anger-management genes included in every modern super-enhancement designer baby gene therapy package.

There were over a dozen fatal car accidents on our nation’s roads last year. The problem is drivers who weren’t enhanced as children and who lack the super-reflexes the rest of us take for granted. This is compounded when they drink before getting on the road, since unenhanced people become impaired by alcohol and their already inferior reflexes deteriorate further. Since the promise of self-driving cars continues to be tied up in regulatory hassles, we can expect many more such needless deaths as long as irresponsible parents continue to consider science “optional”.

And finally, there was a recent outbreak of measles at Disneyland Europa – even though we thought this disease had been eradicated decades ago. Scientists traced the problem to unvaccinated tourists. They further found that all of these unvaccinated individuals were unenhanced. Lacking the cognitive optimization that would help them understand psychoneuroimmunology on an intuitive level, they were easy prey for discredited ideas like “vaccines cause autism”.

So no, super-enhancing your kids isn’t a “personal choice”. It’s your basic duty as a parent and a responsible human being. People in places like India and Neo-Songhai and Venus which suffer from crime and disease make great personal sacrifices to get their children to gene therapy clinics and give them the super-enhancement designer baby gene injection that ensures them a better life. And you start off in a privileged position in America, benefitting from the superenhancement of millions of your fellow citizens, and you think you can just say “No thanks”?

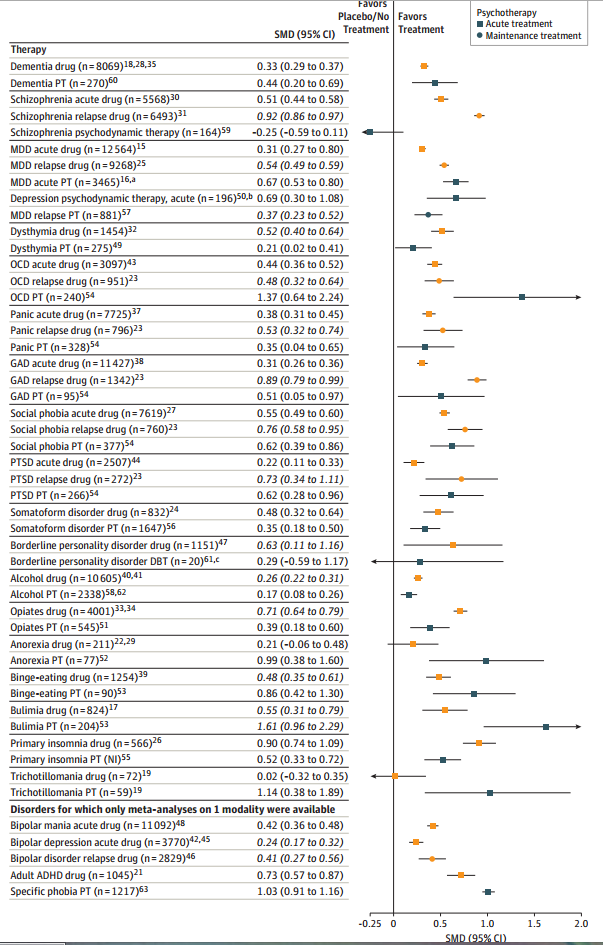
So I don’t want to hear another word from the “but my freedom!” crowd. Unenhanced kids shouldn’t be allowed in school. They shouldn’t be allowed to drive. They shouldn’t be allowed in public places where they can cause problems. And parents who refuse to enhance their children should be put in jail, the same as anyone else whose actions lead to death and suffering. Because not super-enhancing your kids isn’t a “choice”. It’s child abuse.

Mora LeQuivalence is an Assistant Professor of Bioethics at Facebook University. Her latest book, “A Flight Too Far”, argues that the recent Danish experiment with giving children wings is a disgusting offense against the natural order and should be banned worldwide and prosecuted in the International Criminal Court. It is available for 0.02Ƀ on Amazon.com

Related: Transhumanism Is Simplified Humanism, Alicorn’s Alternate Universe Social Justice Series

# The Efficacy Of Everything In Psychiatry In One Graph Plus Several Pages Of Dense But Necessary Explanation

Shamelessly stolen from my hospital’s Journal Club: Huhn et al (2014) graph the Efficacy Of Pharmacotherapy And Psychotherapy For Adult Psychiatric Disorders, and it looks like this:



Before anything else – we kind of have to assume that in each case they’re getting a representative sample of the best drugs/therapies for the disorder. In practice, there is this weird equivalency for most things: most common antidepressants work about equally well, most common antipsychotics work about equally well, et cetera. I don’t know as much about therapy, but I get the impression the same thing goes on there too. So probably it’s not much of a stretch to expect that the efficacy of whatever they studied at least kind of translates to the effectiveness of whatever real treatment you’ll get from your own psychiatrist. At the very least, even lossy and compressed information like this will tell us something.

The effect sizes are mostly around 0.5, with a few much higher and a few much lower. This is common for these sorts of studies. See for example Leucht et al, Putting the efficacy of psychiatric and general medical medication into perspective, which also finds psychiatric effect sizes average around 0.5 and finds this is about equal to average effect sizes in other fields of medicine – thus debunking the popular claim that psychiatry is less effective. Leucht and a few other authors from that piece are also involved in this one, which doesn’t surprise me much.

I do however admit my statistical ignorance in exactly what is going on here. Effect sizes are a good way to compare two unlike domains – for example, I recently noted that leading physicists are about as smart as NBA players are tall. This paper is within that tradition. In fact, if we wanted, we could describe psychiatric medications as about one-sixth as effective as NBA players are tall. This is perfectly honorable. The height of NBA players is a tough bar to live up to.

But I don’t have a good intuitive feel for what it means to use standard mean differences along a non-normally distributed variable – as psychiatric diseases no doubt are. And I’m not sure where they’re even getting their distributions from. When they say schizophrenia meds have an effect size of 0.52, are they talking about the distribution of the general population, with almost everyone near zero and a few schizophrenics way off to the right? Are they talking about the distribution of how schizophrenic particular schizophrenics are, which for all I know might be a bell curve but which is probably very different depending on how you took your schizophrenia sample? I really don’t get this and it’s preventing me from getting a good feeling of exactly how comparable these numbers are to each other.

If we just assume they’re allowed to do what they’re doing, their graph looks about how I would expect it to look. Most psychiatrists always figured that the psychotic disorders were more susceptible to medication and the anxiety disorders to psychotherapy. But three surprises stand out.

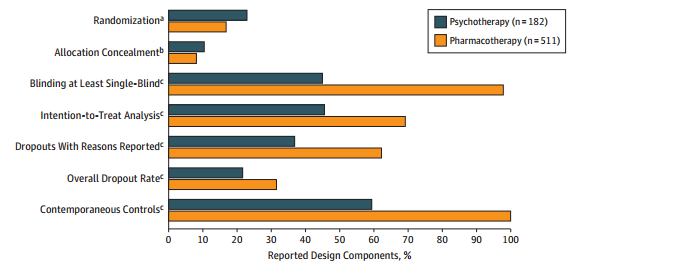
First, this graph shows that drugs are more effective than therapy in treating borderline personality disorder. That’s the opposite of the conventional wisdom, which says that some drugs can decrease impulsiveness in this population but that the definitive treatment has always been Dialectical Behavioral Therapy. But it looks like their borderline psychotherapy “meta-analysis” had a sample size of 20 patients (I would hate to see what the individual studies had!) compared to thousands of patients for most of the believable results. So I wouldn’t place too much faith in this anomaly for now and would continue to recommend psychotherapy for borderlines.

Second, this graph shows that drugs are more effective than therapy for insomnia. Now, we use drugs instead of therapy for insomnia, but conventional wisdom had always been that this was very sad, and there was great therapy available for insomnia if only somebody would provide it. But here the therapy looks mediocre at best. On the other hand, the sample size is “NI”, which I don’t know what it means but doesn’t sound promising. Also, now that anticholinergics probably cause dementia, every single sleeping pill now officially has terrible side effects.

Third, in all conditions drugs seem more effective at preventing relapse than at stopping acute episodes. My “clinical experience,” which is the fancy word doctors use for anecdotal evidence, was exactly the opposite. I now realize I probably faced a lot of selection bias – the patients who do well on their drug and don’t relapse might never see me again. Also, I have a feeling that a lot of the people who come back to me a month later and say “Well, your drug must not have worked, I’ve relapsed again” probably weren’t taking the drug correctly or at all, something which these studies probably enforce better than I can.

In general, the table seems to support psychotherapy being better than drugs for a lot of things. This would not be too surprising if true – their list is heavily tilted to the kinds of conditions therapy works well on – but a caveat is necessary.

The psychotherapy trials were generally of lower quality. Part of this has to do with the culture of psychotherapy research, but more has to do with the underlying territory – giving people “placebo psychotherapy” is more complicated than giving people a sugar pill and a lot of studies don’t bother. Also, in psychotherapy, it tends to be the patient’s therapist recording results more often than corresponding pharmacology studies use the prescriber to record results. That eliminates another layer of blinding.



This has serious ramifications. The study finds that “low-quality psychotherapy trials in general had a higher effect size (SMD = 0.74) than high-quality trials (SMD = 0.22), p < 0.001". Those high effect sizes for psychotherapy aren't looking so good now, are they?

Actually, reread that one more time. Effect sizes for the low quality trials are triple those for the high-quality trials. If you ever wanted proof that it’s way too easy to inflate positive findings if your science isn’t really exceptionally good, there you go.

The most important domain where pharmacotherapy trials are worse than psychotherapy trials is publication bias. The paper suggests that this is because psychotherapy’s lack of sufficient blinding and control groups makes publication results unnecessary. In other words, psychotherapy research isn’t even good enough to have publication bias, because publication bias at least requires you to be rigorous enough to occasionally turn up a negative result to suppress. Ouch.

# Black People Less Likely

[Content warning: Polyamory, race]

I.

The best reporting on social science statistics, like the best reporting in most areas, comes from The Onion:

CAMBRIDGE, MA—A Harvard University study of more than 2,500 middle-income African-American families found that, when compared to other ethnic groups in the same income bracket, blacks were up to 23 percent more likely. “Our data would seem to discredit the notion that black Americans are less likely,” said head researcher Russell Waterstone, noting the study also found that women of African descent were no more or less prone than Latinas. “In fact, over the past several decades, we’ve seen the African-American community nearly triple in probability.” The study noted that, furthermore, Asian-Americans.

I thought of this today because a bunch of people have accosted me about the article There’s A Big Problem With Polyamory That Nobody’s Talking About. “Scott, you’re polyamorous! What do you think of this?”

As per the article, the big problem with polyamorous people is:

…their whiteness. And that standard of whiteness not only erases the experience of people of color; it reflects the actual exclusion of these people in poly life and communities. […]

A white, affluent image that reflects a troubling reality: A 2013 survey of polyamorous people from online groups, mailing lists and forums found that almost 90% of the participants identified as Caucasian. People of color, especially black polyamorists, report feeling “othered” and excluded in poly environments such as meet-ups, with women feeling especially at risk of being objectified and fetishized as an exotic sexual plaything.

“I interviewed a black couple who went to a poly group, and they were definitely preyed upon, in a sense,” said Marla Renee Stewart, Atlanta-based founder of Velvet Lips, a sex education venue.

The article constantly equivocates between “the problem is that polyamory is too white” and “the problem is that the media portrays polyamory as too white”, which is kind of a weird combination of problems to be discussing in a media portrayal. But it seems to eventually settle on a thesis that black people really are strongly underrepresented.

For the record, here is a small sample of other communities where black people are strongly underrepresented:

Runners (3%). Bikers (6%). Furries (2%). Wall Street senior management (2%). Occupy Wall Street protesters (unknown but low, one source says 1.6% but likely an underestimate). BDSM (unknown but low) Tea Party members (1%). American Buddhists (~2%). Bird watchers (4%). Environmentalists (various but universally low). Wikipedia contributors (unknown but low). Atheists (2%). Vegetarian activists (maybe 1-5%). Yoga enthusiasts (unknown but low). College baseball players (5%). Swimmers (2%). Fanfiction readers (2%). Unitarian Universalists (1%).

Can you see what all of these groups have in common?

No. No you can’t. If there’s some hidden factor uniting Wall Street senior management and furries, it is way beyond any of our pay grades.

But what I noticed when I looked up those numbers was that in every case, the people involved have come up with a pat explanation that sounds perfectly plausible right up until you compare it to any other group, at which point it bursts into flames.

For example, Some people explain try to explain declining black interest in baseball by appeal to how some baseball personality made some horribly racist remark. But Donald Sterling continues to be racist as heck, and black people continue to be more than three-quarters of basketball players.

Some people try to explain black people’s underrepresentation on fanfiction websites by saying that many of them have limited access to the Internet. Okay. Except that black people are heavily overrepresented on Twitter, making up double the expected proportion of that site’s population.

Some people try to explain the underrepresentation of blacks in libertarianism and the Tea Party by arguing that these groups’ political beliefs are contrary to black people’s life experiences. But blacks are also underrepresented in groups with precisely the opposite politics. That they make up only 1.6% of visitors to the Occupy Wall Street website is no doubt confounded by who visits websites, but even people who looked at the protests agree that there was a stunning shortage of black faces. I would have liked to get current membership statistics for the US Communist Party, but they weren’t available, so I fudged by looking at the photos of people who “liked” the US Communist Party’s Facebook page. 3% of them were black. Blacks are more likely to endorse environmentalism than whites, but less likely to be involved in the environmentalist movement.

Some people try to explain black people’s underrepresentation on Wall Street by saying Wall Street is racist and intolerant. But Unitarian Universalists are just about the most tolerant people in the world – nobody even knows what they do, just that they’re extremely tolerant when they do it – and black people are in Unitarianism at lower rates than they’re on Wall Street.

And the article on polyamory suggested that maybe polyamorists’ high-flying lifestyle and expensive play parties price out black people. Forget for a moment that I’ve been poly for three years and had no idea this high-flying lifestyle existed and kind of feel like I am missing out. Forget for a moment that as far as I can tell “play parties” are a BDSM term with no relationship to polyamory. In my experience polyamory draws from the same sort of people as atheism, and atheism is very white even though not believing in God doesn’t cost a cent.

This entire genre seems to be a bunch of really silly ad hoc arguments by people who aren’t talking to each other. I would guess most of the underrepresentation of black people in all of these things are for the same couple of reasons.

First, some of these things require some level of affluence – I know I just said that didn’t explain polyamory, but I think it explains some others. For example, bird-watching requires you live somewhere suburban or rural where there are interesting birds, want to waste money on binoculars, and have some free time. Swimming requires you live in an area where the schools or at least the neighborhoods have pools.

Second, Maslow’s Hierarchy Of Needs says you’re not going to do weird things to self-actualize until you feel materially safe and secure. A lot of black people don’t feel like they’re in a position where they can start worrying about where the best bird-watching is at.

Third, the thrive-survive dichotomy says materially insecure people are going to value community and conformity more. Polyamory is still pretty transgressive, and unless you feel very safe or feel sufficiently mobile and atomized that you don’t care what your community thinks about you, you’re not going to feel comfortable making that transgression. Many of these things require leaving the general community to participate in a weird insular subculture, and that requires a sort of lack of preexisting community bonds that I think only comes with the upper middle class.

Fourth, black people might avoid weird nonconformist groups because they’re already on thin enough ice in terms of social acceptance. Being a black person probably already exposes you to enough stigma, without becoming a furry as well.

Fifth, we already know that neighborhoods and churches tend to end up mostly monoracial through a complicated process of aggregating small acts of self-segregation based on slight preferences not to be completely surrounded by people of a different race. It doesn’t seem too unlikely to me that a similar process could act on hobbies and interest groups.

Sixth, even when black people are involved in weird subcultures, they may do them separately from white people, leading white people to think their hobby is almost all white – and leading mostly white academics to miss them in their studies. I once heard about a professor who accused Alcoholics Anonymous of being racist, on the grounds that its membership was almost entirely white. The (white) professor had surveyed AA groups in his (white) neighborhood and asked his (white) friends and (white) grad students to do the same. Meanwhile, when more sober minds (no pun intended) investigated, they found black areas had thriving majority-black AA communities.

Seventh, a lot of groups are stratified by education level. Black people are only about half as likely to have a bachelor’s degree. This matters a lot in areas like atheism that are disproportionately limited to the most educated individuals. Polyamory also falls into this category – the most recent survey found 85% of poly people had a college education, compared to 30% of the general population (!). 30% of poly people had a graduate degree compared to only about 10% of the general population and only about 3% of blacks. There has to be a strong education filter on polyamory to produce those kinds of numbers, and I think that alone is big enough to explain most of the black underrepresentation.

Eighth, people of the same social class tend to cluster, and black people are disproportionately underrepresented among the upper middle class. Most of these fields are dominated by upper middle class people. The nickname for weird self-actualizing upper middle class things is “Stuff White People Like”, and this is not a coincidence. [EDIT: Commenter John Schilling says this better than I – a lot of these groups are about differentiating yourself from a presumedly boring low-status middle class existence, but black people fought hard to get into the middle class, or are still fighting, and are less excited about differentiating themselves from it.]

So I think positing that black people feel “fetishized as an exotic sexual plaything” in the poly community is unnecessary. Black people are underrepresented in the poly community for the same reason they’re underrepresented in everything in the same vague circle as poly. Heck, black people are even underrepresented in the activity of complaining about black people being fetishized as exotic sexual playthings – check out Tumblr’s racial demographics if you don’t believe me.

II.

The eight points above add up to a likelihood that black people will probably be underrepresented in a lot of weird subculturey nonconformist things. This is not a firm law – black people will be overrepresented in a few weird subculturey nonconformist things that are an especially good fit for their culture – but overall I think the rule holds. And that’s a big problem.

A few paragraphs back I mentioned that Occupy Wall Street was had disproportionately few minorities. Here are some other people who like to mention this: Michelle Malkin. The Daily Caller. American Thinker. View From The Right. New York Post. American Renaissance.

All of these sources have something in common, and it’s not a heartfelt concern for equal minority representation.

Likewise, you know who’s got an obsessively large collection of resources on the underrepresentation of minorities in atheism? Conservapedia (Western Atheism And Race, Racial Demographics Of The Richard Dawkins Audience, Richard Dawkins’ Lack Of Appeal To The Asian Woman Audience, etc, etc, not to mention the very classy Richard Dawkins’ Family Fortune And The Slave Trade.)

Here it is easy to see that “you have low minority representation” serves as a stand-in for “you’re racist” serves as a stand-in for “you suck”. So here’s the problem:

In theocracies ruled by the will of God, people will find that God hates weird people who refuse to conform.

In philosopher-kingdoms ruled by pure reason, people will find that pure reason condemns weird people who refuse to conform.

And in enlightened liberal democracies where we “tolerate anything except intolerance”, people will find that weird people who refuse to conform are intolerant.

And if blacks are underrepresented in weird nonconformist groups, and nobody mentions that this is a general principle, that’s making their job way too easy.

So here’s why this article annoys me. In the midst of black underrepresentation in everything in the same ontological category as polyamory, people bring up black underrepresentation in polyamory and suggest it’s because poly people are “objectifying” and “preying on” them, positing that “there’s a problem” with “a standard of whiteness that erases people of color” in the polyamory community.

We know from OKCupid statistics that (mostly monogamous) white men are very reluctant to date black women, but monogamous people don’t have to listen to well-meaning friends going up to them and saying “So, you’re mono, I hear the monogamous community has a racism problem.”

But now I and other polyamorous people are going to have to answer one more round of annoying questions about “You’re polyamorous? Isn’t that a bunch of racist nerdy white dudes?”

# Money, Money, Everywhere, But Not A Cent To Spend

The DSM is written mostly by academics, which is why it gets so excited about distinctions like schizoid personality versus schizotypal personality. If it were written by clinicians, it might better reflect the sort of cases that make it into a hospital.

There would, for example, be an entire chapter on the scourge of ‘My Boyfriend Broke Up With Me’ spectrum disorders. More attention would get paid to the plague of chronic ‘I Got Angry At My Dad And Told Him I Was Going To Kill Myself To Freak Him Out And He Overreacted And Called The Cops And Now Here I Am In Hospital But Honestly I Didn’t Mean It’. Society would finally wake up to the epidemic of ‘I Wanted To Take My Medicine But My Hand Slipped And I Somehow Took The Entire Bottle All At Once Even Though I Would Never Do Something Like Intentionally Overdose’. And the sufferers of ‘This Patient Probably Has Some Kind Of Complicated Neurological Problem But Neurology Is Tired Of Trying To Figure It Out So They Have Declared It To Be Psychiatric’ might at last get some relief.

But the biggest change to the medical lexicon would be the introduction of ‘Poverty NOS’.

I recently got a patient, let’s call him Paul…

(all of my patient stories are vague composites of a bunch of people with details changed to protect privacy)

…who was in hospital after trying to hang himself. He said he was so deep in debt he was never going to get out. He’d been involved in a messy court case, had to hire a lawyer to defend himself, lawyer ended up running to the tune of several thousand dollars. He was a clerk at a clothing store, barely made minimum wage, maxed out his credit cards, then maxed out other credit cards paying off the first credit cards.

He didn’t seem to have major depressive disorder, but when someone comes in admitting to a serious suicide attempt, procedure says he gets committed. He wasn’t thrilled about this, saying if he missed work then he might lose his job and this was just going to make him further behind on his payments, but I checked with my attending and as usual the answer was “admit”.

Something especially bothered me about this case, and after thinking about it I’ve figured out what it is.

It’s not just that the psychiatric hospitalization won’t help and might hurt. That’s pretty common. The ‘My Boyfriend Broke Up With Me’s, the ‘I Got Angry At My Dad’s, unless they have some underlying disorder all of these people get limited value from the psychiatric system and tend to just sit in hospital for a couple of days, go to some group therapy, get asked a hundred times if they’re depressed, then go home. And then they’re still broken up with their boyfriend or still have a terrible relationship with their dad and the same thing’s going to happen again.

In this case, it was that – well, the guy is a minimum wage worker from inner city Detroit. He didn’t tell me exactly how much money this debt was, but from a couple of numbers he mentioned I got the impression it was in the ballpark of $5000. That might not seem like an attempt-suicide level of money to some people, but to this guy with his job the chance of ever paying it off seemed low enough that it wasn’t worth waiting and seeing.

So what bothered me is that psychiatric hospitalization costs about $1,000 a day. Average length of stay for a guy like him might be three to five days. So we were spending $5,000 on his psychiatric hospitalization, which was USELESS, so that we could send him out and he could attempt suicide again because of his $5,000 debt which he has no way of paying off. And probably end up in the hospital a second time, for that matter.

I assume that since he was poor, Medicaid paid his hospital bill. I’m not complaining that the cost of the hospital bill was added to his debt, I’m pretty sure it wasn’t, although in some other cases it would be. I’m complaining that here’s this guy, so desperate for money that he wants to kill himself over it, and he has to sit helplessly as we throw thousands of dollars at getting a parade of expensive doctors and nurses and social workers to talk to him, conclude that yup, his problem is definitely that he’s poor, and then throw him back out. I feel like this fails to be, as the buzzwords say, “patient-centered care”.

Problem is, you don’t have to be an economics PhD to realize that “give $5,000 to anyone who attempts suicide and says they need it” might create some bad incentives.

I have no good solution to this. Offering people who are so poor they want to kill themselves very expensive psychiatric care seems maybe a little better than doing nothing. But it also seems insulting, patronizing, paternalistic, wasteful, and occasionally heartbreaking.

And this is why I can never decide whether to identify as a libertarian or a liberal. On the one hand, top-down institutionalized bureaucracies seem so ridiculously inefficient at solving problems that it’s an outrage and a disaster. On the other hand, there are a lot of problems that really need solving, they don’t seem to have solved themselves yet, and governments are the only entity with enough coordination power to attempt the task.

Solution there, it seems to me, is to create unimpoverishable populaces. I think if we were to implement a Basic Income Guarantee we might save more money in psychiatric care than we think – since we compete with the prison system to be the warehouse for people who can’t make it out in the world and nobody knows what to do with. It might produce some of the same kind of savings as giving the homeless people houses. If I got fired because we’d solved all the problems relating to poverty, and the population of seriously mentally ill people was too small to support the current number of psychiatrists, that would be a pretty neat way to go.

# Drug Testing Welfare Users Is A Sham, But Not For The Reasons You Think

Some people say the War on Drugs is ‘unwinnable’. But there’s actually a foolproof solution that cures drug addiction approximately 100% of the time. That solution is – put people on welfare in Tennessee.

Or at least that is what I am led to believe by articles like Mic’s A Shocking Thing Happened When Tennesee Decided To Drug Test Its Welfare Recipients, which describes said shocking thing as:

1 out of 812 applicants tested positive for drugs. One. Single. Person. Tennessee conservatives suspicious that welfare recipients are a bunch of drug-addicted slackers were proven dead wrong. Big surprise!

After instituting dehumanizing drug-testing requirements to welfare recipients on July 1, 10 people total were flagged for possible drug use and asked to submit to testing. Five others tested negative, and four were rejected after refusing. As Think Progress notes, that means that just 0.12% of all people applying for cash assistance in Tennessee have tested positive for drugs, compared to the 8% who have reported using drugs in the past month among the state’s general population. If you assume the four people who refused were on drugs, it’s still a paltry 0.61%.

In other words, the plan intended to verify right-wing beliefs that welfare recipients are a bunch of drug-addicted slackers looking for a handout has demonstrated exactly the opposite.

The article has 11,000 notes on Tumblr right now, I’ve seen it all over my Facebook feed as well, and the same story has been taken up, with the same editorial line, by a host of other news sources. Jezebel: State Drug Program Busts A Whopping 37 Welfare Applicants. Wall Street Journal: Few Welfare Applicants Caught In Drug Screening Net So Far. New Republic: Red States’ New Tax On The Poor. Daily Kos: Tennessee Just Wasted A Lot Of Money Drug Testing Welfare Recipients. ReverbPress: Another GOP Fail: 0.2% Of Tennessee Welfare Recipients Found To Use Illegal Drugs. Mommyish: Results Of State Drug Testing Prove Gross Assumptions About Welfare Applicants Are Wrong. Washington Post: Scott Walker’s Yellow Politics.

These stories all make the point that we have many stereotypes about the poor, and one such stereotype is that the use lots of drugs, but in fact these sorts of welfare programs find them to use fewer drugs than the general population, and therefore we should stop being so prejudiced.

And if they were found to use only two-thirds, or half as many drugs as the general population, this might indeed be the lesson.

But look at the numbers in the quoted Mic article. Welfare users use only about one percent as many drugs as the general population. Really?

No. Not really at all. According to legitimate research in this area, poor people use as many drugs as anyone else and probably more. The National Household Survey on Drug Abuse found that illegal drug use was slightly higher in families on government assistance (9.6%) than families not on government assistance (6.8%). The National Coalition For The Homeless notes that about 26% of them use drugs, which is about 2.5x as high as the general population. I crunched some data I have from the hospital I work at, and it shows that poor people (defined as people who get health insurance through an aid program) have moderately higher rates of drug use related problems than the general population. So these articles are reporting a drug use rate in the Tennessee population about one percent of that ever reported in any comparable poor population anywhere else.

Kate from Gruntled and Hinged brings up another curious inconsistency. The false positive rate for drug tests is – well, it depends on the test procedure, but it’s usually at least 1%. So if every single welfare user in Tennessee was 100% clean, we would still expect between 1% to 5% positive drug tests. Instead, they got 0.12% positive drug tests. This isn’t just suspiciously good, it’s impossibly good.

So what’s going on here?

Before I explain, here’s a collage of the stock photos displayed above some of those news stories I linked to.



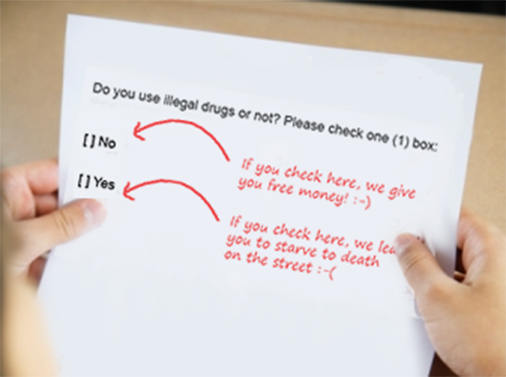
I now have a picture on my website called urine\_collage.png

If you’re familiar with the state of the American media, you won’t be surprised to learn that urine was not involved in the ovewhelming majority of this program’s drug tests.

So how did they test people for drugs?

They gave them a written test, where the test question was basically “do you use illegal drugs or not?” You can see the exact procedure on the sidebar here.

And lo and behold, the overwhelming majority of people answered that they didn’t.



A more accurate stock photo they could have used

Now the numbers make sense. It’s not that only 0.2% of welfare recipients use drugs. All this tells us, if anything, is that 0.2% of welfare recipients are on so many drugs they can’t figure out how to check “NO” on a form.

Why would the government do something like this? As best I can tell, the plan was originally to give everyone urine checks, but in Florida the courts decided that urine-checking people without prior suspicion was unconstitutional. The Republicans were pretty attached to their “drug test welfare recipients” plan and didn’t want to look like they were wimps who backed down just because of one little court case, so they decided to give people the written test in the hopes of having prior suspicion for the people who said yes. Sure, it made no sense, but they could still tell their constituents they were drug testing those welfare recipients, and in principle they’d won an important victory. Or something.

Which raises another interesting question – how did Florida’s urine-based program do before the courts struck it down?

According to the media, abysmally. MSNBC: Drug Testing Welfare Recipients Looks Even Worse, “[Florida Governor] Scott’s policy was an embarrassing flop. Only about 2 percent of applicants tested positive, and Florida actually lost money”. TBO: Welfare Drug Testing Yields 2% Positive Results, “Newton said that’s proof the drug-testing program is based on a stereotype, not hard facts.” ATTN: Why Drug Testing Poor People Is A Waste Of Time And Money, “Florida tested welfare recipients for four months before its drug test mandate was thrown out by the courts. Only 2.6 percent of welfare recipients tested positive. The rest of the Florida’s population use drugs at a rate of 8 percent. So, again, welfare recipients used drugs less than everyone else.”

Now we’re merely at one-quarter of the drug use rate people with good methodologies find. Improvement!

So I looked up exactly how this works. Apparently welfare recipients were asked to pay for their own drug tests, and would be reimbursed if the results came back negative. 7000 welfare users did this, but 1600 declined to do so – numbers that were not mentioned in most of the pieces above.

Opponents of the program say that maybe those 1600 people could not find drug testing centers near them, or couldn’t afford to pay for the tests even with the promise of reimbursement later, or something like that. I am sure that some of them did indeed decline for reasons like those.

But also, people on welfare don’t have very much money [citation needed]. If I were a welfare recipient, and they were going to drug test me and not reimburse me if I came out positive, and I was on drugs, I would decline the hell out of that test.

Suppose that the poor in Florida use drugs at the same rate as the poor in various studies and surveys – about 10%. We have 8600 welfare recipients, so we would expect 860 drug users. Of the 7000 who agreed to testing, we know that 2.5% are drug users – that’s 175 people. That in turn would suggest that of the 1600 who refused testing, about 685 were drug users – 40% or so. That would imply that about 80% of drug users versus about 12% of nonusers refused testing.

These numbers seem pretty reasonable to me. Most welfare users want to keep their benefits, so the majority will agree to testing, but a few will inevitably fall through the cracks because they can’t reach a testing center or because they have moral objections to the tests. On the other hand, clued-in drug users will realize that for them, testing means a major inconvenience and monetary charge without any likely corresponding gain. So we would expect drug users to decline testing at a higher rate than nonusers. In order to use the Florida data to say that welfare recipients in general use drugs at a rate of 2%, we would need to assume that drug users were no more likely to refuse drug testing than nonusers, even though the testing rewarded non-use with money but punished use with a loss of money.

(note that there are some different numbers in different places for Florida. I assume that these represent different years, stages of testing, parts of Florida, etc, but I’m not sure. The only one that is seriously different from what I’m saying above is the one that says “only 1% of people declined testing”. After some search, I’m pretty sure that’s referring to that only 1% of people made appointments for testing, then cancelled later. But I am less confident in the Florida numbers than in the analysis of Tennessee)

So the Florida numbers are consistent with welfare recipients using drugs less, more, or the same amount as the general population.

So I have a question for you guys.

How come Brian Williams is being dragged over the coals for lying in the media, but everyone who publishes these kinds of articles gets off scot-free?

If I understand correctly, Williams said that his helicopter got shot at when he was in Iraq, but in reality he was just in a helicopter in Iraq at the same time as some other helicopter nearby was getting shot at. This is obviously stretching the truth, but it seems to me it could have been worse. No important policy decisions are going to hinge upon exactly which helicopter Brian Williams was in. And he didn’t get it infinitely wrong – for example, there was, in some sense, a war in Iraq.

On the other hand, discussions of how many poor people use drugs is pretty important for all sorts of policy questions, and these people completely dropped the ball. So why does nobody get reprimanded for this kind of thing?

You might argue that Brian Williams’ actions were obviously malicious and deceitful, but that screwing up drug numbers is an excusable mistake. I say it’s exactly the opposite. Brian Williams did exactly what I unfortunately do all the time – unthinkingly tell a story the much cooler way it should have happened, the way it happened in my head – rather than the way it actually did happen (my colleagues elsewhere in the psychiatry blogosphere go further and call this “normal brain function”).

On the other hand, I have more trouble imagining a situation in which I would accept the claim “only 0.1% of poor people use drugs, which is barely one percent of the rate in the general population” without wanting to do a little more research to see if it is true. If your reporters are capable of making this mistake honestly, get better reporters.

But I’m not sure it’s honest. A lot of these sources admit they took their story from a Think Progress piece on the issue. Think Progress does mention that the tests are a sham, although only in one sentence that is easy to miss. Either the secondary reporters didn’t read Think Progress thoroughly, or they consciously decided not to mention it.

But even if it was an honest mistake, I still have trouble excusing their arrogance. I mean look at that Jezebel article. The writer says this proves that people who think welfare recipients use drugs “consider ‘facts’ troublesome” and that their “entire social philosophy boils down to ‘Ew, poor people.'”

You’re saying that’s not as bad as a helicopter-related embellishment?

Yes, okay, drug testing welfare applicants is in fact probably a bad idea. It’s a bad idea because the courts have banned doing it in a way more effective than asking them politely if they use drugs or not, but it was a bad idea even before that. It’s a bad idea because drug tests have frequent false positives, but it’s a bad idea even without that. It’s a bad idea because quitting drugs is really hard and denying people benefits isn’t going to help.

But if, in the service of proving this to be a bad idea, you decide it’s acceptable to fudge the numbers to make your point, horrible things happen. First, you contribute to a culture of telling lies and lose the opportunity to protest when the other side does it. Second, you make it harder to trust you on anything else.

But most important, tell one lie and the truth is forever after your enemy. I recently argued that we need to reform suboxone prescribing laws, because it’s the best anti-addiction medicine we’ve got and right now poor people can’t access it. . Why should anyone listen to me now? They can just answer “Actually, that would be a waste of money. As per an article I read in Jezebel, pretty much no poor person has ever been addicted to drugs.” Then the laws don’t get reformed and people die.

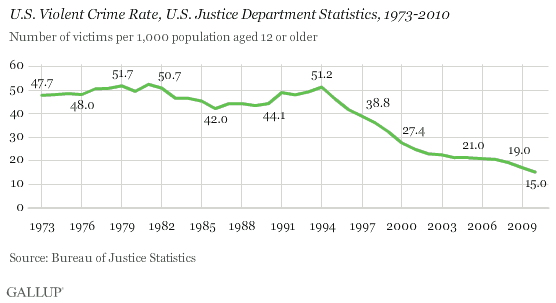
# How Likely Are Multifactorial Trends?

Vox recently wrote about 16 Theories For Why Crime Plummeted In The US.

Their story is based on a report by the Brennan Center For Justice, which I haven’t read, so I’m hesitant to critique it too much. The little I got off of Vox I don’t like. For example, if I understand correctly they’re arguing that the lead-crime connection is overblown because although lead was banned in the 1970s (thus affecting people who reached peak crime-committing age in the 1990s), the decline in crime continued even into the 2000s. But lead stays in the environment a long time, there’s still a lot of work to be done eliminating various sources of lead, and so blood lead levels continue to decline. That makes their argument ring a little hollow.

But I want to talk about a more meta-level point.

The analysis ends up concluding that there is no “smoking gun” and crime probably declined because of a bunch of reasons coming together. For example, they say that “up to 12 percent of the drop in property crime during the 1990s was due to the rise in incarceration, but it was probably more like 6 percent”, and “up to ten percent of the drop in crime in the 1990s was caused by hiring more police.” The general picture I get is that there were about ten different factors, each explaining ten percent of the decline.



Imagine two different perspectives on this.

First, a learned professor says “Oh yes, the public always wants to hear about how one big exciting thing caused the decline in crime, but that kind of thinking is unsophisticated. Something as complicated as crime is governed by many factors, and you certainly wouldn’t expect one big knockout change to lower it to this degree. Like everything else, it’s probably a combination of different things that came together, each accounting for a small percent of the variance.”

Second, someone counterargues: “If ten different factors caused the decline in crime, that would require that ten different things suddenly changed direction, all at the same time in 1994. That’s a pretty big coincidence. In fact, let’s reductio ad absurdum this. Imagine it was ten million different factors, each accounting for one ten-millionth of the decline. But that seems stupid. For example, since there are only about ten million criminals in the US, we could structure this as one factor per criminal. Imagine that, in 1994, each of America’s ten million criminals independently and coincidentally had a major life change that made crime seem less attractive. That’s ridiculous. But in that case, any other explanation based on ten million factors should seem ridiculous. And if we give a heavy credibility penalty to a story with ten million factors, we should give some credibility penalty to a story with ten factors.”

The second person seems to me to have a strong argument, which makes me think Vox and the Brennan Center’s model where ten different trends each explain about ten percent of the decline is unlikely.

I feel like somebody has already thought about this and there’s an entire literature I’m missing, but Google is failing me (badly – this was my first search result). Can somebody point me to it? Are there ways to calculate how much less likely a ten-factor explanation is than a one-factor explanation?

[EDIT: Yes, there’s the trivial case where all ten factors are correlated, for example they all have to do with an improving economy. I’m talking about the non-boring version of the question.]

[EDIT2: I might have subconsciously absorbed this thought process from Stefan Schubert]

# Did Falling Testosterone Affect Falling Crime?

There are already too many proposed causes for the secular decline in crime, but I can’t resist suggesting one more. A couple of months ago Nydwracu asked me whether it could be related to the secular decline in testosterone. The answer turns out to be “Maybe”.

This secular decline in testosterone is pretty dramatic. Our best source is A Population-Level Decline In Serum Testosterone Levels In American Men, which finds that from 1987 to 2004, average testosterone declined from 501 ng/dl to 391 ng/dl, with an even more dramatic decline in bioavailable levels of the hormone. That’s about minus 1% per year.

No one knows exactly why this is happening. Some people blame increasing obesity and decreasing tobacco use (wait? Smoking increases testosterone levels? THOSE TV COWBOYS WERE RIGHT ALL ALONG!). Other people have tried to adjust for these and found they don’t explain the entire effect, leading to a host of other theories. Recent scrutiny has focused on the role of feminizing chemicals in the water supply, probably a combination of industrial pollutants and discarded medications; the worst-affected areas are marked by an epidemic of transsexual fish (really).

(A quick aside – since these chemicals are gender-bending fish, frogs, and various other animals, could they be responsible for transgender in humans? This theory seems to still be in crackpot territory, but I don’t know why. Research shows that male-assigned-at-birth children exposed to diethylstilbestrol in the womb are more likely to become transgender than the general population. Other than that, there just seems to be one unpublished paper on the subject. Get to work, scientists!)

Annnnnyway, testosterone has been found to correlate a bit with violent crime. In a study of 692 male criminals, Dabbs et al found that those in prison for more violent crimes had higher testosterone than those in prison for nonviolent offenses. It’s hard to say exactly how much higher because they report their testosterone in a different way that doesn’t correlate to anyone else’s – I think part of it is that it’s salivary rather than serum testosterone but it’s still confusing even after I adjust for that. If we use relative rather than absolute, they do mention that 66% of inmates in the upper third of testosterone levels committed violent crimes compared to 46% in the lower third. High-T inmates were twice as likely to be in for murder as low-T inmates. Interestingly, testosterone was the highest risk factor for sex crimes, such as child molestation and (especially) rape – high-T inmates were four times as likely to be in for rape as low-T inmates. On the other hand, low-T inmates were about twice as likely to be in prison for drug offenses.

This “which criminals are worse” study is obviously not as good as an “are high-T people more likely to be criminals at all” study, but I can’t fin any of those with a good sample size. You can read a review of the research here.

According to the population decline study, testosterone levels declined about 110 ng/dL in 15 years. They don’t give me a standard deviation, but from this site I get one a bit less than 200. So testosterone declines by one standard deviation about 25 years? That means that a person in the top third of testosterone levels today would have been in the bottom third fifty years ago. Which – and I realize I’m doing all sorts of horrible things here to cover up my lack of actually useful data – if we extrapolate wildly from the results of these studies, we could sort of justify murder halving in about fifty years by falling testosterone alone.

The first problem with this is that we can’t really use data on prison inmates as representative of the population.

The second problem is that murder has halved in way less than fifty years. It seems to have halved between 1994 and 2004.

The third problem is that crime didn’t start falling until the early 1990s, but testosterone was falling since at least 1987 and probably earlier. This site, which doesn’t cite sources, says testosterone was higher in the 1940s, though they might be confusing that with “in men born in the 1940s, as studied in the 1980s”, which is of uncertain significance. Sperm count has been declining since the 30s, according to an article called Sperm Quality & Quantity Declining, Mounting Evidence Suggests

(it looks like somebody was not quite as virtuous as this Twitter user).

The fourth problem is that there’s contradictory evidence about whether testosterone is even falling at all, according to a a study that looked at the faces of Major League Baseball players of the past 120 years. This sort of makes sense – face width-height ratio is affected by testosterone (one reason women’s faces look different than men’s) and baseball players had standardized photographs taken of them for that time period. They find that, at least based on the face ratios, testosterone was increasing during that period, which would be interesting if it didn’t contradict everybody else. As it is, I suspect it just means baseball players were differently representative of the general population. For example, if baseball requires high testosterone, and scouts became better at selecting the highest-testosterone people over that period, that’d do it. Or if the nature of baseball changed to more of a “power game” rather than a “finesse game” (I think some people have said this) that’d do it too. Or if all baseball players suddenly started taking powerful testosterone-analogue chemicals at some point…hmmmmmmmmm…On the other hand – literally on the hand – we have the digit ratios of Lithuanians over 120 years. Someone in 1880 measured the length of Lithuanians’ fingers – which can be a proxy for testosterone levels – and then the experiment was repeated recently and the results compared. It did find the expected increase in testosterone, though no word on whether that was throughout the entire period or just concentrated in the past couple of years. So this sort of turned out to be a non-problem.

The fifth problem is that crime is dropping in women at the same rate as in men – women never really committed that many crimes, but now they’re committing fewer. Women do have some testosterone, so it’s possible that declining testosterone could affect female violence as well, but it wouldn’t be the first thing I expected. Also, I’m not sure if there are any secular trends in female testosterone levels, though I’d be fascinated to see data.

So overall while I like the approach of this hypothesis, I don’t think it gets the time window right. It would be a nice way to explain a gradual fifty-year decrease in violent crime starting in the 50s and continuing to the present day. Instead, we have a big spike in the 50s and a big drop in the 90s, which were not particularly abnormal in terms of testosterone decline.

This doesn’t really make sense to me. If testosterone is declining, it should cause a decrease in crime. One might argue that testosterone levels have been steadily operating behind the scenes causing very long term declines while other things account for the more visible short-term trends, but that seems like a cop-out.

I’d like to see studies comparing testosterone levels in violent criminals (both male and female) to those in the general population.

Also, we have cemeteries full of millions of dead people from every era of history, all carefully marked with what age they were when they died. Somebody needs to dig some of them up and measure their digit ratios – I assume you can still measure the digit ratio of bones, the overall length is still there. Then we can have a good answer for whether testosterone levels in men (and women) have been declining over time, when it started, and whether it’s been picking up recently. If it has been, the chance that it hasn’t had an important effect on our society worth exploring is pretty much nil.

I know, just once I want to get through an entire blog post without a call for disinterring the dead, but this is important.

# Pharma Virumque

Going around the psychiatry blogosphere recently: this segment by John Oliver about doctors who take pharmaceutical company money:

I will resist the urge to geek out about its minor medical errors1 in favor of clarifying something more important.

The impression you’re supposed to get from this piece is a shady looking man handing you a briefcase full of cash and whispering “Hey, here’s $10,000 for you if you prescribe unnecessary medication.” The implication is the doctors who do this are awful and if you were in medicine you would have no trouble resisting this temptation.

In reality, pharma companies have figured out that some people have ethical qualms – “evil cannot possibly understand good” only works in movies – and adjusted their strategies accordingly.

We’ll start with a simple one. Imagine you’re a doctor, and your staff are complaining because the staff at every other doctor’s office has been getting these incredible free lunches every day – the video says drug companies aren’t supposed to give, like, Zagat-rated steakhouse lunches, but there’s still a lot of room between “Zagat-rated” and “Way better than the peanut butter and jelly sandwich you bring from home”. The nurses are grumbling and threatening to revolt and asking if you really appreciate them.

A drug company representative offers to provide your office with free lunches a couple of times a week.

You say “It would be really annoying to actually use the phrase ‘there’s no such thing as a free lunch’ here, so I will just ask what the catch is.”

They say “No catch. We don’t require you to ever prescribe any of our drugs. We don’t require you to listen to our presentation. We don’t even require you to read our promotional literature. Just accept our offer.”

You say “Why are you doing this?”

They say “Because every time you eat one of our lunches, you’ll associate the ice cold taste of Coca-Cola and the sweet warm chewy chocolate chip cookies with our company, and you’ll get positive feelings about it, and maybe those positive feelings will influence your prescription habits.”

You say “I think I’m a good enough doctor not to prescribe a drug solely because I get lunch from their company.”

They say “Look. We all know that most antidepressants are about equally effective. Sure, we split hairs and talk about how one has more anticholinergic side effects so it’s bad for patients with cholinergic sensitivity, and another has more chance of weird visual disturbances, but how often does someone come into your office and announce ‘Hey, I’m depressed, and also I have cholinergic sensitivity, but I LOVE weird visual disturbances!’? Although there are a few cases where one drug’s clearly a better choice than another, most of the time you’re about equally balanced between two or three options, and you just pick one at random. So maybe instead of picking one at random, you’ll pick the one you associate with delicious food. And if you do, so what? Nobody’s harmed. You would have just flipped a coin anyway.”

You say “I’d rather flip a coin than feel like I’m being pressured by what I had for lunch.”

They say “Look, you secretly worry anyway that you sometimes prescribe Effexor because the name makes it sound effective, or Paxil because the name makes it sound peaceful.”

You say “Wait, you can read my thoughts?”

They say “We’re a pharmaceutical company. Of course we can read your thoughts. Look. You already know that the mostly-meaningless choice of which of several equally effective drugs you prescribe is influenced by a bunch of silly marketing factors beyond your control. Why not add one more?”

“But -”

“Come to the Dark Side! We literally have cookies!”

Still not tempting enough for you?

Okay, imagine this. You’re a doctor and one of your patients comes in with incurable chronic pain that’s ruining their life. You try the normal medications on it and nothing works very well. There’s a high-tech next-generation medication available that you think is a good fit for your patient’s disease, but it’s not covered by their insurance and there’s no way the patient can afford it. You have to tell this guy that there’s nothing you can do for him.

Then a drug company representative comes to you bearing a big box of free samples. By “free samples” I mean hundreds of pills, enough to help the patient for the better part of a year – and maybe at the end of that time you’ll get another box of free samples. The drug rep doesn’t want you to sign your life away. She’s giving them for free, no obligation, maybe just listen to a sixty second speech on how to prescribe them safely and effectively (she wouldn’t want to give them to someone who won’t prescribe them effectively!) Are you really so fundamentalist in your approach to medical ethics that you won’t listen to a drug rep for sixty seconds in order to save a patient’s quality of life?

Most doctors – even the ethical ones who would refuse the briefcase full of cash – take the offer. This practice has come under increasing scrutiny recently. Some of the complaints are kind of dumb, but one very valid one is that a lot of the times what happens is you start off by giving the patient 100 days of free sample or something, then the free sample runs out, they’re fixated on that particular medication because it’s the one that worked for them, and they find some costly way to continue the (more expensive new) medication – instead of the two of you working harder to find some older less expensive medication that works equally well. A few drug companies have “fixed” this by giving out cards for “prescription programs” that solve some of the problems with free samples. These are even harder to resist, and they’re also given out by attractive drug reps who just want to tell you a few important facts about the drug before giving it to you.

Still not tempting enough for you?

Fine, then imagine this. You’re a doctor who really believes in a particular drug and is trying to convince the medical community to use more of it. For example, a couple weeks ago I wrote an article on suboxone saying it was one of the best medications for opiate abuse and I wish the medical community would pay more attention and prescribe it more often.

I wrote that article for free as a public service because I think that drug saves lives. But imagine that the company that makes suboxone approached me afterwards and said “Hey, you seem to have an important message to spread. Why don’t we sponsor you to go around the country for a week or two telling it to other doctors at medical conferences? We’ll get you first-class flights, put you up in five-star hotels, and give you a $10,000 stipend.”

I say “Wait a second, that sounds like taking pharmaceutical company money, and taking pharmaceutical company money is evil.”

They say “Look. You were trying to promote suboxone to people already. You were just doing a bad job because you were limited to one little blog. The more suboxone-promoting you do, the more doctors know about this drug – which you yourself have said is life-saving – and so the more lives get saved. If you’re willing to promote suboxone ineffectively for free, why not promote suboxone effectively for $10,000 plus nice hotels?”

I say “I’m still kind of uncomfortable with this.”

They say “Okay, well, it’s not our fault if hundreds of people die of drug overdoses because their doctors didn’t know suboxone was an option.”

You’re probably going to ask if I’ve ever accepted any of these offers. The boring truth is that I haven’t had to consider them because I’m a resident and residents are lower than dirt and the pharmaceutical companies know this and they don’t waste time trying to cozy up to us.

I have tasted the forbidden fruit only once, and it was my attending’s fault. She told us that there was a big dinner being planned for the entire psychiatric community of our city. The goal was to get doctors to meet nurses to meet therapists to meet social workers in one place so we could all get to know each other and talk about changes we could make to the system. It was very important that we attend, or else the nurses and therapists and social workers would think that the doctors were too snooty to interact with them and didn’t care about changing the system. Oh, and by the way the dinner was sponsored by PANEXA (here used in place of the real drug because I don’t want to get in trouble for calling them out) but there wouldn’t be any promotional material or pressure to prescribe PANEXA, honest, no sirree.

This was a tempting offer precisely because it was such a good idea. Everyone in the local psychiatric community deals with each other frequently, but we’d mostly never met before. I know them as the voice on the other side of the phone saying “No, no beds are available in our facility” or as the person who refuses to fax me my patient’s past medical history because the patient is too catatonic to sign a consent form. None of us are ever entirely sure what the others are doing, sometimes there are bad feelings, and it was reasonable to hope that maybe if we all met each other and socialized things would get a little smoother.

So we all meet at this restaurant, and immediately World War III breaks out. It’s like “Hi, I’m Mary, the clerk at Blue Sky Mental Health.” “MARY?! YOU’RE THE ONE WHO DIDN’T FAX ME THOSE RECORDS I NEEDED TWO MONTHS AGO! MY PATIENT WENT A WEEK ON THE WRONG DRUGS BECAUSE OF THAT!”

“Hi, I’m Dr. Alexander, I work at the inpatient unit in Our Lady Of An Undisclosed Location Hospital…” “WE HAD A PATIENT COME FROM THERE TWO WEEKS AGO AND HE ASSAULTED A STAFF MEMBER. IF YOU’RE A REAL HOSPITAL WHY CAN’T YOU DO PROPER VIOLENCE ASSESSMENTS?”

It turned out that the nurses hated the social workers for making them wait on the phone forever in order to get a straight answer. The social workers hated the nurses for always calling them up when they were busy about things and expecting an answer RIGHT NOW. The social workers hated the doctors for giving patients one measly prescription, then handing the case over to them to fix all of the impossible problems in the patient’s life. The doctors hated the social workers, because when we give patients one measly prescription and then hand the case over to the social workers to fix all of the patient’s impossible problems, sometimes the impossible problems don’t get fixed.

Anyway, in the midst of all of this, there was one guy who was staying completely calm, talking nicely to everybody, helping people see each other’s sides of the issue, just a really serene well-adjusted guy. I escaped over to his table and asked him who he was and why he was here.

“Oh,” he said “I’m a paranoid schizophrenic currently on PANEXA.”

Of course he was.

Then we all broke off into our own groups and got some incredible Italian food.2

What I’m saying is, pharmaceutical companies are sneaky.

Footnotes

1: By which I mean “succumb to the urge to geek out about its minor medical errors, but in the footnotes”.

The video says that a “horrifying example” of pharmaceutical company overreach was how AstraZeneca took Seroquel, “an antipsychotic with dangerous side effects” and marketed it to doctors for depression, sleep, and dementia, adding “You can’t just give people dangerous drugs and see what happens!”

But actually, lots of studies have shown Seroquel is effective for depression, lots of guidelines suggest Seroquel as a backup depression treatment, and doctors have been (correctly) prescribing it for such for a long time. Doctors also very commonly prescribe it for sleep and dementia; I think is less evidence-based, but it’d be a lie to say it wasn’t common as dirt or that it didn’t work for these things (safety is the problem).

So what was happening was that AstraZeneca was promoting Seroquel for the things it was actually being used for, as opposed to the thing the FDA said it was supposed to be used for. Doctors are allowed to use drugs for whatever they want based on their own analysis and their best judgment, but pharmaceutical companies are only allowed to promote it for the FDA-approved indication, which at that point was psychosis and bipolar depression.

The reason the FDA hadn’t approved Seroquel for depression wasn’t because it was a bad idea. It was because in order to get the FDA to approve anything for anything, you must perform the appropriate ritual of putting a zillion dollars into a big pile, then burning it as a sacrifice to the Bureaucracy Gods. AstraZeneca had performed the ritual for bipolar and psychosis, but was still in the process of performing it a third time for depression. Once they finished, the FDA approved it as an adjunctive medication for depression, but also fined them hundreds of millions of dollars because they had advertised it for depression – merely based on evidence and clinical practice – before the FDA had told them they were allowed to.

This is still not the whole story, because best clinical practice says to only use Seroquel as a third- or fourth-line antidepressant after some others have failed, and in conjunction with another medication. If AstraZeneca was advocating to use it for depression first-line on its own, this would have been a genuine overstep and something to get upset about.

(research and clinical practice say to use it for sleep and dementia approximately never, but there is enough wiggle room in that “approximately” for doctors to drive a bus through, and they do.)

This is still not the whole story, because The Last Psychiatrist thinks the way the FDA’s handled the Seroquel indication, and the subsequent culture of prescribing that grew up based on that indication, is stupid.

The other minor medical error in the video is much simpler. Oliver mocks Wellbutrin’s claim to be “the happy, healthy skinny drug” saying that “the only happy, healthy, skinny drug is amphetamine”. But Wellbutrin is actually amphetamine-based – its full chemical name is 3-chloro-N-tert-butyl-β-ketoamphetamine – and it shares a mechanism of action with amphetamines, which is why some of its effects are similar as well. So Oliver’s joke was a lot more accurate and a lot less funny then he thought.

2: Then later, and contrary to the promises I received, they gave us a presentation on PANEXA anyway.

The schizophrenic guy worked for one of the local psychiatric community services groups doing community outreach. I never did figure out whether he was there as a coincidence or whether the pharmaceutical company had arranged to have him there. I suspect the latter but I have no proof.

# California Meetups

The following Slate Star Codex meetups are planned for the next week or so:

Berkeley, 3/1 at 2:00 PM at Indian Flavor Express, 2548 Bancroft Way, Berkeley

Stanford, 3/3 at 4:00 PM in Old Union, Room 200, 450 Serra Mall, Stanford

There’s still a little bit of planning going on around the second one, so watch for possible updates.

I have been subpoenaed by a court in Michigan to testify about one of my patients. I am trying to convince them that I am on vacation and it’s unfair to make me come back on such short notice. If they start threatening me with legal penalties for not showing up, I’ve got to cut my vacation short and these meetups will not happen (any legal experts here who can tell me a way to get out of this will be duly appreciated). Again, watch for possible updates.

# Early Intervention: You \*Might\* Get What You Pay For

I find myself caught between the genetics community – which takes it as a given that childhood experiences and education have a very limited role in shaping life outcomes – and the psychiatric community, which takes it as a given that childhood experiences and education are crucial in shaping life outcomes. Both sides have their favorite studies to cite supporting their positions. I’ve already talked about the genetics studies, so I thought I’d bring up a recent particularly good study from the other side.

Dodge et al’s Impact Of Early Intervention On Psychopathology, Crime, And Well-Being At Age 25 is published in last month’s American Journal Of Psychiatry. Gratifyingly, it is a randomized controlled trial. Ten thousand kindergarteners in disadvantaged areas were screened for “conduct problems” until they found about 900 who looked like they were at high risk. 445 were randomly selected for the intervention. Another 446 stayed in the control group. The intervention was a bunch of extra classes and ‘enrichment programs’ from elementary school (age 5) all the way through high school (age 16). The study mentions “social skills friendship groups”, “guided parent child interaction sessions”, “tutoring in reading”, “parent-youth groups on topics of adolescent development, alcohol, tobacco, and drugs”, “youth forums on vocational opportunities”, and “Oysterman’s School-To-Job possible selves intervention aimed at examining emerging identity”.

All of these sound so pretentious that I would have loved to be able to report that they had no effect, but in fact the opposite was true. When they caught up with these kids at age 25, the intervention group was found to have an odds ratio of around 0.6 to 0.7 of having developed various psychiatric disorders the study was testing for, including antisocial personality disorder, ADHD, depression, or anxiety. They had odds ratios around 0.7 of developing drug and alcohol abuse problems by various measures. They reported less risky sexual behavior, less domestic abuse, and fewer violent crimes. All of this was significant at the p < 0.05 level, and some of it was significant at much higher levels like p = 0.001 or below. Subgroup analysis found the data were very similar when you restricted the analysis to various subgroups like boys, girls, whites, blacks, highest-risk, lowest-risk, and by study site (it was a multi-site study). As best I can tell there were not an equal number of anaylses they did that came up negative that they covered up.

The apparent conclusion is that intensive interventions can change children’s outcomes and personalities in important ways ten years down the road, even regarding things believed to be highly genetic like antisocial personality disorder.

A few weak attempts to rebut this. First, there were some things that study didn’t do that one might have expected it to. It didn’t change graduation rates or employment rates. The apparent decrease in domestic violence was mediated entirely by the intervention group being less likely to have relationships (!) – the rate of domestic violence among people in relationships was the same. There was no effect on health. There was no effect on self-reported satisfaction with their parents’ parenting. There were (nonsignificantly) higher death rates and incarceration rates in the intervention group than the control group.

So if I wanted to be maximally mean to the study, I could say that whatever it’s doing to violent crime and drug use has to be compatible with a (nonsignificantly) raised incarceration rate, and whatever it’s doing to drug use and risky sexual behavior and criminality has to be compatible with a (nonsignificantly) raised death rate. This suggests the possibility of an attack based on their endpoints being screwy, though I’m not sure what form such an attack could take. One could argue that since many of their outcomes were based on self-report surveys maybe the kids who had been through all of the enrichment programs had grown to like the study people and had a stronger demand effect to say that they were doing great. But a lot of the survey data was backed up by court records confirming fewer drug and violence convictions. So that doesn’t really work.

If you’re less interested in the pure science of individual differences and more interested in policy, one fact that I forgot to mention was that this program cost $60,000 per kid. The paper points out that this is the same cost as a year or two of incarceration, so if it really changes children’s life outcomes and makes tham less antisocial even that hefty price tag might be justified (although again, remember that it didn’t affect employment or incarceration when checked directly).

If you’re looking for an optimistic spin on that number, they freely admit they have no idea which part of their gigantic ten year intervention program produced the positive effects. It could be that all the youth forums and enrichment programs and friendship groups and so on had zero effect, and the entire benefit came from the “Oysterman’s School-To-Job possible selves intervention aimed at examining emerging identity”. And maybe that’s a piece of paper that can be copied on a copy machine for ten cents a sheet. All this suggests is that at least some part of the ten-year, $60,000 intervention did something.

If you’re looking for a pessimistic spin on that number, consider. Every so often I see things that claim to have completely shifted children from the most high-risk of high-risk groups to upstanding successful members of society by giving them a year or preschool, or a couple of after-school lessons, or something like that. And these studies always boast that they did it with only $1000 or $5000 or some number like that, so it’s nice and cost effective. So far, the studies I have seen like this have been wrong. And so far I have not been surprised, because we already spend between $100,000 to $200,000 per child on education and various social programs. If someone ever found a social program that really worked for $1,000, the first thing we would want to do is tar and feather everyone currently in our bureaucracy of social programs, for being so incompetent that changing their $200,000 in spending to $201,000 in spending (with the extra $1000 going to someone besides them) could completely revolutionize life outcomes.

This study seems more in line with everything else. By going from $200,000 to $260,000, we can slightly push a few things in a positive direction a little bit more, maybe. From a scientific view, it’s pretty interesting. From a policy view, it’s nothing to write home about.

# California Meetups (Rererevised)

The following Slate Star Codex meetups are planned for the next week or so:

Berkeley, Sunday 3/1 at 2:00 PM at Indian Flavor Express, 2548 Bancroft Way, Berkeley

San Jose, Monday 3/2 at 7:30 PM at David Friedman’s house, 3806 Williams Rd, San Jose

Googleplex, Tuesday 3/3 at 1:00 PM at the picnic tables between CL2 and CL5, maybe. Googlers can get more information at go/ssc-meetup

Stanford, Tuesday 3/3 at 6:00 PM, at the picnic tables outside Tresidder Union, 459 Lagunita Dr, Stanford

across the street from Old Union, Room 200, 450 Serra Mall, Stanford, at the picnic tables outside Tresidder

EVERYONE is welcome to come to Berkeley and San Jose, even if you don’t read this much, don’t comment, whatever. Googleplex is limited to Google employees, unless you have another way to get in (NOT to be interpreted as a challenge!). Stanford is mostly intended for Stanford students, but if you’re in the area and really want to come I’m sure we can fit a few others in.

# A Series Of Unprincipled Exceptions

The more consistently one attempts to adhere to an ideology, the more one's sanity becomes a series of unprincipled exceptions.

— graaaaaagh (@graaaaaagh) February 5, 2015

Meeting with a large group of effective altruists can be a philosophically disconcerting experience, and my recent meetup with Stanford Effective Altruist Club was no exception.

Buck forced me to pay attention to an argument I’ve been carefully avoiding. Most people intuitively believe that animals have non-zero moral value; it’s worse to torture a dog than to not do that. Most people also believe their moral value is some function of the animal’s complexity and intelligence which leaves them less morally important than humans but not infinitely less morally important than humans. Most people then conclude that probably the welfare of animals is moderately important in the same way the welfare of various other demographic groups like elderly people or Norwegians is moderately important – one more thing to plug into the moral calculus.

In reality it’s pretty hard to come up with way of valuing animals that makes this work. If it takes a thousand chickens to have the moral weight of one human, the importance of chicken suffering alone is probably within an order of magnitude of all human suffering. You would need to set your weights remarkably precisely for the values of global animal suffering and global human suffering to even be in the same ballpark. Barring that amazing coincidence, either you shouldn’t care about animals at all or they should totally swamp every other concern. Most of what would seem like otherwise reasonable premises suggest the “totally swamp every other concern” branch.

So if you’re actually an effective altruist, the sort of person who wants your do-gooding to do the most good per unit resource, you should be focusing entirely on animal-related charities and totally ignoring humans (except insofar as humans actions affect animals; worrying about x-risk is probably still okay).

I acknowledged the argument was very convincing, but told Buck that I was basically going to safe-word out of that level of utilitarian reasoning, for the sake of my sanity.

Buck pointed out that this shouldn’t be too scary, given that many utilitarians have already had to go through a similar process. Peter Singer talks about widening circles of concern. First you move from total selfishness to an understanding that your friends and family are people just like you and need to be treated with respect and understanding. Then you go from just your friends and family to everyone in your community. Then you go from just your community to all humanity. Then you go from just humanity to all animals.

By the time most people figure out what they’re doing they already accept at least friends, family, and community. But going from “just my community” to “also foreigners” is a difficult step that’s kind of at the heart of the effective altruism movement. In the same way that allowing animals into the circle of concern totally pushes out the value of all humans, allowing starving Third World people into the circle of concern totally pushes out most First World charities like art museums and school music programs and holiday food drives. This is a scary discovery and most people shy away from it. Effective altruists are the people who are selected for not having shied away from it. So why shy away from doing the same with animals?

It’s a good question. After thinking about it for a while, I think my answer is that I never actually completed the process of widening my circles of concern and neither has anybody else, and because I’m thinking about this one in an abstract intellectual way I’m imagining actually completing it, which would be much scarier than the incomplete things I’ve done before.

Like, although I acknowledge my friends and family as important people whom I should try to help, in reality I don’t treat them as quite as important as myself. If my brother asked me for money, I’d lend it to him, but I wouldn’t give him exactly half my money no-strings-attached on the grounds that he is exactly as important to me as I am.

Likewise, although I acknowledge strangers as important people whom I should try to help, in reality I don’t treat them as quite as important as my friends. We all raised a lot of money to help Multi when she was in a bad situation, but there are thousands of other people in the same exact same bad situation and we’re not putting nearly as much effort into them.

You can try to justify this in terms of “well, I know myself better than I know my brother, and I know Multi better than I know strangers, so I’m more effective at helping me and Multi, so I’m just rationally doing the things that would have the most impact”. But I think if I bothered to dream up some thought experiment where that wasn’t true, I would prefer to help me and Multi to my brother and random strangers even after that factor had been controlled away.

This doesn’t come as a surprise to me and I’m not sorry. But…well…I guess my worry about the animal charity thing wasn’t that I was inconsistent, so much as that I was being meta-inconsistent; that is, I didn’t even have a consistent set of rules for deciding whether I was going to want to be consistent or not.

And now I think I might have a consistent policy of allowing some of my resources into each new circle of concern while also holding back the rest of it for the sake of my sanity. Thus my endorsement of GiveWell’s principle that you should donate at least 10% of your income to charity, but then feel okay about not donating more if you don’t want to. I am allowed to balance resources devoted to sanity versus morality and decide how much of what I have I want to send into each new circle of concern – without denying that the circle exists.

I think that armed with this idea I am willing to accept Buck’s argument about animal welfare being more important than human welfare, insofar as this means I should donate some resources to animal welfare without necessarily having to give up caring about human welfare completely. I don’t think I can make a principled defense of doing this. But I think I can claim I’m being unprincipled in a meta-consistent and effectively sanity-protecting way.

# A Cascade Of Dunbar Numbers

Never doubt that a small group of thoughtful, committed citizens can become a large group of thoughtless, committed citizens.

— Member Of Species (@MemberOfSpecies) January 27, 2015

Back when I used to play at micronations, we had grand dreams of creating a fully-fleshed out simulated country with hundreds of citizens and dozens of Byzantine government departments and industries. For a while, these dreams seemed attainable. Most of the people we advertised to were really interested, and we went from one or two founding members to about a dozen people very quickly.

Then we stopped. For the past fifteen years or so, the population of Shireroth remained fixed at between ten to twenty people. It wasn’t always the same ten or twenty – some new people entered, some old people left – but it always stayed within that window.

Ten years ago we got really tired of this and spent a decent amount of money on an advertising campaign. It was a good campaign, and we got maybe thirty or forty people to apply for citizenship at once. We thought this was going to be so awesome – the size of the simulation quintupling within a week. Instead, the new people just kind of wandered around aimlessly, bumping into things, never quite figuring out what was going on, never quite taking an active role in anything or showing any initiative, finally showing up less and less until two months later the population was back down to about a dozen people.

Our Ministry of Immigration and Naturalization, which was in charge of recruiting new people, kept beating its head up against the same sort of idea with about the same results. I thought this was dumb, so I manuevered my way into the Ministry and shifted all of our resources from recruitment to retention – setting up mentoring programs for new citizens, writing a bunch of guides for what they were supposed to do, trying to befriend them and integrate them into our social scene. The result was that our new immigrants maybe stayed three months instead of two months before evaporating.

We were only ever able to find one successful method of recruiting new micronationalists: wait for random people around the world to learn about micronations, start their own projects without any impact or influence from us, integrate those projects into the larger community, and then if those projects failed after a year or two sometimes their members would join us instead. This sort of worked, but never enough to get us more than ten or twenty people. And those other projects, the ones they made, very rarely had more than ten or twenty people either, despite hordes of people who said they thought micronations were interesting and agreed to participate for a few months. It was a weird problem, and one I was never able to solve before I left the hobby.

I was thinking about this recently because of some people’s complaints about the Bay Area rationalist community.

I have always had universally positive experiences with the Bay Area rationalist community. I arrived in about 2012 when it was still smallish, and I got to know the group involved and hang out with fascinating people like Eliezer Yudkowsky and Anna Salamon and Michael Vassar and learn a lot from them.

The complaints I heard were something like this. A lot of people join rationalist communities in Missouri or Italy or Australia or some far-flung place like that and really like them and feel like they have a great tight-knit community. They decide that if a small rationalist community is good, a ten-times-larger rationalist community will be ten times as good. Then they get here and everyone’s already doing their own thing, and it turns out Eliezer and Anna and Michael do not have enough time to personally interact with every single person in San Francisco on an intimate basis, and they kind of hang around the edges of the community not really knowing what to do or being connected to its general rhythm. It doesn’t help that a lot of people (DESPITE MY BEST EFFORTS) keep thinking that rationality will turn them into supermen and get upset that they’ve been in the Bay Area like an entire month and are not yet Elon Musk.

And I was thinking of both of these things when I met Samo in Oakland and he told me about his research. It’s kind of complicated, but when I asked him “Is this basically hypothesizing a cascade of Dunbar numbers?” he told me I wasn’t far off.

Dunbar’s number, remember, is the theory of anthropoligst Robin Dunbar that humans have a hard-coded optimal group size. He found that most primates’ brain size correlates with the average size of their social groups. If you extrapolate the connection to humans, you would expect humans to have a social group of about a hundred fifty people (other estimates find anywhere between 100 and 250). Various researchers and amateurs have claimed to corroborate this in every context from hunter-gatherer bands to corporations to MMORPG guilds (1, 2), sometimes more believably than others.

Samo said his own research had found several of these discontinuities – I think he mentioned about 12 people, about 150 people, about 1,000 people, and about 90,000 people as the first few. He tried to tie this in to government forms like “family”, “clan”, “tribe”, and “city-state”. I think the idea was supposed to be that we’re naturally very good at dealing with groups of family size, our brains will grudgingly allow us to deal with groups of up to clan size, with the low-hanging fruit in social technology we can stay together in groups of tribe size, with some higher-hanging fruit in groups of up to city-state size, and so on. When you get to the largest organizations in the world (China and the Catholic Church, if you’re wondering), you have to pull out all the stops – a pyramid of subdivisions of subdivisions of subdivisions, each choosing representatives to the next-highest level which then chooses its own representatives and so on until you reach one guy at the top. He suggested that if my old micronation or the Bay Area rationalist community want to get past their growth limitations, they should try something similar – which is good advice, except that in fifteen years of trying to simulate a country it did occur to us to have provinces and that didn’t seem to help much.

I am not sure how much I buy this theory. Having a systematizing disposition, it is really tempting for me to start carving up reality at these lines – startups with < 12 people, small businesses with < 150, et cetera - or to start noting that most of the college clubs I've seen have like a dozen core members, no matter how big the college is. But I find myself engaging in certain thought patterns that I have tagged as "trying too hard to fit data to a theory". At University College Cork the debating club often got twenty or thirty people; and if someone told me the natural way to slice businesess was <5 people, <30 people, et cetera, I could kind of see that too.  
  
And also there are so many other factors - whether a group is online vs. meatspace, whether it's living together versus meeting once every few weeks, whether it's strongly enforced (like a company or military unit) or voluntary (like a social club), whether we expect its members to also have other relationships somewhere else (members of MMORPG guilds presumably have real-life friends and coworkers as well, but members of hunter-gatherer tribes might know literally nobody except their cotribesmen). Even if there were interesting limits to the size of cohesive social groups, I would expect these factors to make them different in different domains - the size restrictions governing the growth of online micronations should be different from the size restrictions governing the growth of startups, even though in practice they're often around the same size.  
  
But I do notice that the Bay Area rationalist community is probably around 150 people and having trouble growing, and the size of my micronation was around 12 people and had trouble growing. And that after ten years in micronations, the rule that no matter what happened we would never get more than a dozen or so people seemed like a law of nature. And that I just checked after two years outside the community, and it looks like one new person came in to replace me and no one else, and they’re still stuck around twelve. So there does seem to be this very real community failure mode where everyone is socially saturated with each other and new people feel like they can’t break in. I’m less confident it always happens at a specific number or set of numbers across all domains of community – but who knows?

# Effective Altruists: Not As Mentally Ill As You Think

During my recent meetings with effective altruist groups here, I kept hearing the theory that effective altruism selects for people with mental disorders. The theory is that people with a lot of depression, anxiety, and self-hatred turn to effective altruism as (optimistically) a way to prove that they are good and valuable or (pessimistically) a form of self-harm in which they enact their belief that they deserve nothing and other people are more worthy.

And whenever this got brought up at meetings, people giggled, probably because they were thinking of good examples. I can’t deny there’s a lot of anecdotal evidence here (hi Ozy!). But when I look into it, it seems totally false.

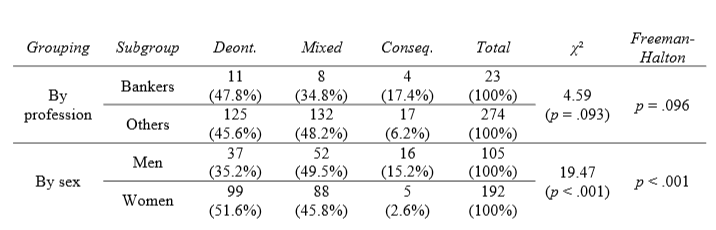
My source was the 2014 Less Wrong survey data, which asked respondents whether they self-identified as effective altruists and whether they participated in effective altruist groups and meetups. Using that question, I separated the respondents into 758 non-effective-altruists and 422 effective altruists. The survey had also asked people whether they had been diagnosed with various mental illnesses, so I checked the rates in both groups. Including self-diagnosis there were no particular results; when I limited it to professionally diagnosed illnesses things got a little more interesting.

Effective altruists had about the same levels of anxiety disorders and obsessive-compulsive disorder as non-EA Less Wrongers. However, they had slightly higher levels of depression (22% vs. 17%) which was barely significant (p = 0.04) due to a large sample size. They also had more autism (8.5% vs. 5%) which was also significant (p = 0.02).

I expected this to be mediated by a tendency for autistic people to be more consequentialist and consequentalists to be more EA, and both these things were true to some degree, but even when I limited the analysis to all consequentialists, effective altruists still had more autism. Further, autistic people seemed to donate a higher percent of their income to charity than neurotypical people or people with other mental illnesses even separated from effective altruist status – that is, even among people none of whom were effective altruists, the autistic people seemed to donate more (effect not always significant) even though they generally had lower incomes.

I conclude that effective altruists are not unusually self-hating or scrupulous, but that they may be a little more autistic, and the reason why isn’t the obvious one.

A caveat, by way of presenting another interesting result. Rusch (2015) (h/t @bechhof) studied whether bankers were more consequentialist (in this case, more likely to give consequentialist answers to the Trolley Problem and Fat Man Problem) than nonbankers. He found that they were. But then he checked for confounders and found the result was entirely an artifact of men being more consequentialist than women and bankers being predominantly male.



This is pretty astounding – men are almost six times as consequentialist as women!

On the other hand, in both my Less Wrong data in general and the effective altruist subgroup, men and women don’t vary much in consequentialismishness. Either Rusch’s data is wrong, or there’s a strong filter that acts to get only consequentialists into Less Wrong regardless of gender, or LW converts women to consequentialism (without further converting men).

Interestingly, effective altruists were much more consequentialist than non-effective-altruist LWers – 80% versus 50%. They also had more women than the non-effective-altruists. So it looks like LW filters for consequentalists so strongly it gets an even balance of consequentialist men and consequentialist women, and past that stage, filtering further for consequentialism doesn’t change gender balance much.

This points out a limitation of my statistics above. All it shows is that effective altruists don’t differ from other rationalists in levels of mental illness. It’s possible and indeed likely that both effective altruists and rationalists differ from the general population in all kinds of ways. It’s even possible that self-hate and scrupulosity drive people into the rationality movement in general, although I can’t imagine why that would be. It’s just that they don’t seem to have any extra power to make people effective altruists once they’re there.

# Too Good To Be True

Related To: You Might Get What You Pay For, Do Life Hacks Ever Reach Fixation?

Here are three interesting psychological studies:

1. Kirschenbaum, Malett, and Humphrey gave students a three month course in how to make monthly plans, then followed them up a year later to see how well their grades were doing. The students who made the plans got an average GPA of 3.3 compared to the students who didn’t getting 2.5. They concluded that plan-making skills are academically important, and that benefits persist at least one year after completion of the plan-making course.

2. Aronson asked Stanford students to write a letter to a middle school “pen pal” urging them to adopt a “growth mindset”. Since this was a psych study, it was all lies and there was no pen pal; the study examined whether writing a letter urging growth mindset made the students themselves have a growth mindset and whether this improved grades. Three months later, the students who wrote the letter had higher GPAs.

3. Oaten and Chang helped undergraduates set up an 8-week time management program involving schedules and diaries. The students who participated not only had better time-management, they also studied more, smoked less, drank less alcohol, exercised more, ate a better diet, spent less money, rated their emotions as better, missed fewer appointments, and were less likely to leave dishes in the sink (really!).

So, remember a couple of weeks ago when I wrote about some psychiatrists conducting a really big (n =~ 1000) study about an early intervention program for troubled youth? A program that cost $58,000 per person and lasted ten years?

And remember how, although it was deemed a success, it was deemed a success because it had modest effects on a couple of outcomes, without improving the really big ones like school retention, employment, or incarceration?

So on the one hand, having a short discussion about making monthly plans will boost your GPA almost a whole point a year later. On the other hand, ten years of private tutoring and pretty much every social service known to mankind will do next to nothing.

This suggests a dilemma: either psychological research sucks or everything else sucks.

I mean, you tell your social engineer “Here’s ten thousand dollars and a thousand hours of class time per pupil per year, go teach our kids stuff,” and they try their hardest.

And then some researcher comes along, performs a quick experimental manipulation (the pen pal one probably took 30 minutes and 30 cents) and dramatically improves outcomes over what the social engineer was able to do on her own.

Then one gets the impression that the social engineer was not using their $10,000 and 1K hours very wisely.

And if it were just the one example, then we could say Carol Dweck or Roy Baumeister or whoever is a genius, the rest of us couldn’t have been expected to come up with that, now that we know we’ll reform the system. But these results have been coming in several times a year for decades. If we’ve been adopting all of them, why aren’t people much better in every way? If we haven’t been adopting them, why not?

My money is on the other branch of the dilemma. The reason the $58,000 study got so much less impressive results is that it was run by medical professionals to medical standards, meaning it only showed the effects that were really there. The reason psychology gets such impressive results is…

Okay. I have only skimmed these three studies, so I don’t want to make it sound like I’m definitively crushing them. But here are some worrying things I notice.

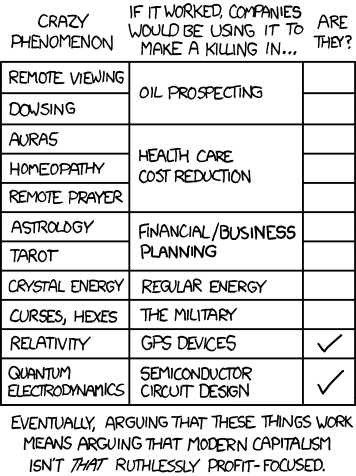
The first study results are actually limited to a small subgroup with I think a single-digit number of students per cell.

The second study results work only on a complex statistical manipulation and disappear when you do basic correlation; further, although the manipulation is supposed to work by increasing trait growth mindset, the correlation between trait growth mindset and academic achievement when correlated directly is actually negative across all variables and in some cases significantly so (!)

The third study is really about stress-related behaviors during an examination period, meaning that all they showed was that people who as part of their time management course were forced to study in a carefully scheduled way for a term show less stress-related behavior during the term exam period, which makes more sense as they probably studied more earlier, had less studying left to do during the exam period, had more free time, and were less stressed. This context was dropped by the popular science press, turning the study into proof that good time management in general always produced all of these effects.

And although there are several other studies I have not been able to find equally worrying flaws with, if they report massive long-term gains from seemingly minor interventions, I expect they’re there and I just missed them.

Basically, you remember this chart?



Source: xkcd

In “Crazy Phenomenon”, add “any large and persistent effect from social psychology”. In “If it worked…” add “education, rehab, and mental health”. In “Are They?”, add “not nearly as much as I would expect”.

# Book Review: Willpower

I.

Willpower: Rediscovering the Greatest Human Strength by Roy Baumeister and John Tierney attracted me with the following pitch: there are only two quantities in psychology that have been robustly linked to a broad range of important life outcomes. One is IQ and isn’t changeable. The other is willpower and is easily changeable. Therefore, study willlpower.

I expected the book to center around Baumeister’s groundbreaking experiments in “ego depletion” – where people who are forced to expend willpower on one task have less willpower left over for future tasks – and “ego repletion”, where people who have been depleted of willpower get it back after taking some glucose. I’d been left kind of confused by competing claims about those studies and I hoped this book would fill out my knowledge of them and settle my confusion.

Instead, it spent a couple of chapters mentioning their existence and praising them as revolutionary, and then got deep enough into Pop Science Self-Help Book Mode to be almost a self-parody. Chapter three is called “The To-Do List From God To Drew Carey” and illustrates why to-do lists are a great idea with anecdotes from Drew Carey’s life and the Bible. Chapter four is the same but with Eliot Spitzer. Chapter six is David Blaine, Chapter seven is H.M. Stanley, Chapter eight is Eric Clapton, and Chapter ten is Oprah. All of these people apparently have important lessons about willpower to teach us, of which the interesting ones are:

– Willpower is a limited resource that is depleted by use and restored by glucose

– Having to make too many small decisions in a day causes “decision fatigue”, leading you to be exhausted and make bad decisions.

– Using willpower a lot strengthens your willpower, allowing you to be more effective.

– People with more willpower do much better in life; for example, in the famous “marshmallow test”, children who were able to resist eating a marshmallow for a few minutes in order to get two marshmallows had better life outcomes twenty years later.

– Careful quantification of goals, setting precommitments, and publicizing your success or failure to other people helps you stick to resolutions. (If you are wondering exactly how to do this, this might be a good time to mention there’s a new ad on the sidebar for Beeminder, a company that manages this for you with some neat evidence-based tricks.)

– Religious people have more willpower than non-religious people for some reason. But you have to believe if you want to get this benefit – you can’t just hang out at church and go through the motions in order to reap the willpower gains.

– Chinese-Americans do better than white Americans on willpower tests from toddlerhood toward adulthood. This is the most likely reason Chinese people outperform whites in the real world, and in fact although average white and Chinese IQ are pretty similar, Chinese people can break into “elite” professions at a lower IQ threshold than whites because their increased self-control compensates. The book admits this may be partly genetic, but also attributes some of it to Chinese parents teaching their children discipline and setting hard goals, which they contrast with white parents who tell their kids to “have fun” and “have high self-esteem” and “be self-directed”. Obviously in order to believe this result you’d have to believe parenting styles can affect children, which this book takes on faith but which is on shaky ground.

– Amount of self-control does not affect body weight (!) or success at dieting (!!) very much.

Overall there were some interesting findings in here, even if I found the pop sci tone a little bit over-the-top after a while.

II.

But I was disappointed. The reason I got this book is that there’s a big debate going on over “willpower”, “ego depletion”, and their younger cousins “growth mindset” and “grit”. All of these are slightly different constructs, but they’re all measures of stick-to-it-ness, and all of their proponents make grand claims about how small interventions to make people have them will bring those people success at school, work, and life. On the other hand, there are a lot of other people who think this whole area is a load of bunk.

For example, Mischel’s marshmallow test started all this off by “proving” that children who were able to delay gratification longest had higher SAT scores, higher parent-rated competence, and better coping skills. But now the test is under fire as people question whether it just shows that people from good environments learn to be more trusting and so more likely to believe the researcher’s promise of extra rewards for delayed gratification. Other people ask if it just shows that kids who are smart enough to think of good strategies to distract themselves are also smart enough to get good SAT scores and all of the other positive correlates of high IQ. Still other people point out the very low sample size, the reverse correlation in other subgroups, and an apparent failure to replicate. Mischel fires back in an Atlantic article where he says of course he took these things into account, that the test was done on a homogenous upper-class population.

This book just says the marshmallow test proves willpower is important, and leaves it at that.

Or how about the idea that glucose is the limited resource that willpower depletes? Robert Kurzban very correctly points out that the metabolic math doesn’t come close to adding up – we know how much glucose things in the body use, and these short little willpower tasks aren’t really going to affect blood glucose levels at all. Also, turns out that if you rinse your mouth out with a tasty glucose solution, you get just the same amount of ego replenishment even though none of the glucose actually entered your body. And for that matter, how come I can’t get infinite willpower just by snacking while I work? How come M&Ms don’t work as a poor man’s Adderall?

Kurzban goes further and says he doesn’t believe in willpower as a limited resource at all. He notes that even when you’ve stopped studying because you’re too ego-depleted and exhausted to make yourself go on any further, if I offer you a million dollars to study another hour then you’ll do it. Guess that resource wasn’t so depleted after all. He proposes a different model of willpower, where it’s your brain’s way of nagging you about the opportunity cost of your actions – “you’ve been studying three whole hours, don’t you think you could use that lobe of your brain for something else now?” But this strikes me as ridiculous – my brain is very concerned that it has better things to do than study, but is perfectly happy with me playing Civilization IV: Fall From Heaven forever or simply lying in bed doing nothing?

Finally, Carol Dweck finds that willpower is only depletable if you think it is, which sounds like exactly the sort of thing Carol Dweck would find. If Carol Dweck ever became an oncologist, we would have to revise all the medical textbooks to say that people only get cancer if they think they will.

There is a meta-analysis of about a hundred studies said Baumeister was basically right about everything. On the other hand, Baumeister’s theory failed what sounds like a formal replication. So it’s complicated.

I really want to know what willpower is. It seems like one of the big challenges of my life; I never have enough willpower for everything I want to do, and I’d at least like to have a theory of what I’m up against. This book did not give me enough information to navigate the controversy. Instead, it totally denied there was any controversy and spent chapter after chapter on cute pieces of trivia interspersed with stories about Drew Carey and Oprah.

III.

So let me end this with some thoughts that a good explanation of willpower should take into account.

First, mental willpower seems a lot like physical willpower – by which I mean our ability to push through exhaustion to keep exercising. Kurzban complains that willpower can’t be a limited resource, because even after it’s all depleted you can still force yourself to keep going for a big enough reward. But the same is true of exercise. I’ll get exhausted and stop running after a certain number of miles, but if you offer me $1 million to run another one I’ll probably make it. And the studies showing that rinsing your mouth out with glucose gives you the same willpower boost as actually drinking is identical to the results of similar studies measuring exercise duration.

So it’s probably worth asking what exactly causes exercise fatigue. This is less firmly known than I expected, but it seems to be a combination of decreasing levels of inputs (especially glycogen stores in the muscles), buildup of toxic metabolic waste (including heat, lactic acid, etc), and cellular damage. Exercise long enough and your muscles need time to replenish their stores, clear away waste, and repair themselves.

Second, mental willpower seems to do something a lot like budgeting. Money, despite being pretty much the classic example of a limited resource, shares some of the features of willpower that Kurzban mocks. A few days ago a pipe burst, my house flooded, and I may have to spend several thousand dollars unflooding it. This means I will have no money for the next forever and then some, and if you ask me to meet you at a fancy restaurant I’ll probably refuse on financial grounds. But if my mother is on her deathbed and her final wish is to see me one last time, I will find the money to get an expensive flight to California. How’s that any different from me not having enough willpower to keep studying until you offer me a million dollars to do so?

Likewise, if tomorrow my boss offers me a $5000 bonus to be handed out next month, I’ll probably relax my budgetary constraints and start buying myself nice things even before I get the paycheck. How’s that any different from the body reacting to sugar when it’s in the mouth but not the bloodstream?

Finally, the question that I as a psychiatrist find most interesting – how come drugs can change willpower so dramatically? The guy who can’t concentrate on a project for more than five minutes straight will pull a whole week of all-nighters when he’s on Adderall or modafinil. Those certainly aren’t increasing blood glucose, so what’s up?

The model that makes the most sense to me is of a stupid default system running on short-term reinforcement learning, plus an evolutionarily novel (and therefore poorly implemented) executive system that can overrule the default. The executive system’s overrule isn’t a simple veto, but a constant action, the same way holding your hand high in the air for a long period requires constant action by your muscles. This effort is metabolically costly in the same way that using muscles is metabolically costly, and so your body runs a general-purpose budgeting function on it that convinces you to turn it off before it overheats. Given enough incentive, you can let it overheat, but then it’s going to be damaged and need to repair itself for a few days before you can use it effectively again. This seems to fit all the evidence except the drugs, which I interpret as acting on the default system so that you don’t need to bring in the executive planner. I freely admit this is sort of cheating.

Overall I recommend Willpower if you want a quick and fun survey of a bunch of loosely connected psychology topics, but not if you want a deep and balanced exploration into the literature of anything.

# Answer to Job

(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 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.”

“HMMMMMMMM…” 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

Then 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 @simulacrumbs)

— Scott Alexander (@slatestarcodex) March 13, 2015

# List Of Passages I Highlighted In My Copy Of “Willpower”

[content note: dieting]

Warning: I have not checked any of these claims for truth, I was generally not impressed with the skepticism level displayed in this book, and I highlighted the passages I found most surprising or counterintuitive. So these are quotations, not endorsements. As Ashleigh Brilliant says, “My sources are unreliable, but their information is fascinating.”

A good way to appreciate the Zeigarnik effect is to listen to a randomly chosen song and shut it off halfway through. The song is then likely to run through your mind on its own, at odd intervals. If you get to the end of the song, the mind checks it off, so to speak. If you stop it in the middle, however, the mind treats the song as unfinished business.

[Examination-related words] popped more frequently into the mind of the group who had been told about the exam but hadn’t made plans to study for it. No such effect was seen among the students who’d made a study plan. Even though they, too, had been reminded of the exam, their minds had apparently been cleared by the act of writing down a plan.

In 1995, Tierney did a semiscientific survey of a New York phenomenon: the huge number of intelligent and attractive people who complained that it was impossible to find a romantic partner. Manhattan had the highest percentage of single people in any country in America except for an insland in Hawaii originally settled as a leper colony. What was keeping New Yorkers apart? Tierney surveyed a sampling of personal ads in the city magazines of Boston, Baltimore, Chicago, Los Angeles, and New York. He found that singles in the biggest city, New York, not only had the most choice but were the pickiest in listing the attributes of their desired partners. The average personal ad in New York maganize listed 5.7 criteria required in a partner, significantly more than second place Chicago’s average (4.1 criteria) and about twice the average for the other three cities.

Bob Boice looked into the writing habits of young professors and tracked them to see how they fared. Some of these professors would collect information until they were ready and then write a manuscript in a burst of intsense energy. Others plodded along at a steadier pace, trying to write a page or two every day. When Boice followed up on the group some years later, he foud that their paths had diverged sharply. The page-a-day folks had done well and generally gotten tenure. The so-called “binge writers” had fared far less well, and many of them had had their careers cut short.

As early as the 1920s, researchers reported that students who spent more time in Sunday school scored higher on laboratory tests of self-discipline. Religiously devout children were rated low on impulsiveness by both parents and teachers…But psychologists have found that people who attend religious services for extrinsic reasons, like wanting to impress others or make social connections, don’t have the same high level of self-control as the true believers.

A clear difference between Chinese and American toddlers appears when they’re asked to override their natural impulses. In one test, for instance, the toddlers are shown a series of pictures and instructed to say “day” whenever they see the moon and “night” whenever they see the sun. In other tests, the toddlers try to restrain themselves to a whisper when they’re excited, and play a version of Simon Says. The Chinese four-year-olds generally perform better on these tests than Americans of the same age. The Chinese toddlers’ self-control might be due in part to genes. There’s evidence that the genetic factors associated with ADHD are much rare in Chinese children than in American children. But the cultural traditions undoubtedly play a role as well. Asian-Americans make up only 4% of the US population but account for a quarter of the student body of elite universities like Stanford, Cornell, and Columbia. They’re more likely to get a college degree than any other ethnic group, and they go on to earn salaries that are 25% above the American norm. Their success has led to the popular notion that Asians are more intelligent, but that’s not how James Flynn explains their achievements. After carefully reviewing IQ studies, Flynn concludes that the scores of Chinese-Americans and Japanese-Americans are very similar to Americans of European descent. If anything their IQs are slightly lower, though they do show up more at both the upper and lower extremes. The big difference is they make better use of their intelligence. People working in elite professions like physicians, scientists, and accountants generally have an IQ above a certain threshold. For white Americans, that threshold is IQ 110, but Chinese Americans manage to get the same elite jobs with an IQ of only 103.

When asked how parents could contribute to childrens’ academic success, the mothers who had emigrated from China most frequently mentioned setting high goals, enforcing tough standards, and requiring children to do extra homework. Meanwhile, the native-born mothers of European ancestry were determined not to put too much pressure on children They most frequently mentioned the importance of not overemphasizing academic success, of stressing the child’s social development, and of promoting the idea that “learning is fun” and “not something you work at”. Another of their chief concerns was promoting the child’s self-esteem, a concept of just about no interest to the Chinese mothers in the study.

[In his preliminary marshmallow-test-style experiments in Trinidad], Mischel stumbled upon a bigger and more meaningful effect. Children who had a father in the home were far more willing than others to choose the delayed reward. Most of the racial and ethnic variation could be explained by this difference, because the Indian children tended to live with both parents whereas a fair number of the African children lived with a single mother. These findings, which were published in 1958, didn’t attract much attention at the time or in ensuing decades, since it was dangerous to one’s career to suggest that there might be drawbacks to single-parent homes…One possible explanation is that children in one-parent homes start off with a genetic disadvantage in self-control. After all, if the father (or mother) has run off and abandoned the family, he may have genes favoring impulsive behavior. Some researchers have attempted to correct this by looking at children who were raised by single parents because the father was absent for other reasons (like being stationed overseas, or dying at a young age). Predictably, the results were in between. These children showed some deficits, but their problemes were not as large as those of the children whose fathers had voluntarily left the home.”

Nurture Assumption said the opposite (children whose fathers left for valid reasons were totally indistinguishable from children with two parents), I believed them until I coincidentally found the contrary evidence, and I’m still angry at them for this.

When fat lab rats are put on a controlled diet for the first time, they’ll lose weight. But if they’re then allowed to eat freely again, they’ll gradually fatten up, and if they’re put on another diet it will take longer to lose the weight this time. Then, once they again go off the diet, they’ll regain the weight more quickly than the last time. By the third or fourth time they go through this boom-and-bust cycle, the dieting ceases to work – the extra weight stays even though they’re consuming fewer calories.

If true this would require a pretty high-level uncoupling of calories and weight gain.

An English bookmaker, the William Hill agency, has a standing offer to bet against anyone who makes a plan to lose weight. The bookmaker, which offers odds of up to 50 to 1, lets the bettors set their own target of how much weight to lose in how much time. It seems crazy for a bookmaker to let bettors not only set the terms of the wager but also control its outcome – it’s like letting a runner bet on beating a target time he sets himself. Yet despite these advantages, the bettors lose 80% of the time.”

I tried to find these people’s website to see how the numbers worked, but it wouldn’t let me in because I’m in the US. Brits might want to check this out.

Dieters have a fixed target in mind, and when they exceed it for any reason – [like being told to taste test a milkshake as part of an experiment] they regard their diet as blown for the day. So they think what the hell, I might as well enjoy myself for the day and the resulting binge often puts on far more weight than the original lapse.

People who weigh themselves every day are far more effective at preventing their weight from creeping back up.

Some of the [experimental subjects] were shown a meal from Applebees consisting of chicken salad and a Pepsi; others were shown an identical meal with some added crackers prominently labeled “Trans Fat Free”. The people were so entranced by the crackers’ virtuous label that their estimate for the meal with crackers was lower than that for the same meal without the crackers!

A new form of the Conjunction Fallacy?

To mimic the human studies [on ego depletion], experimenters depleted the willpower of one group of dogs by having each dog obey “sit” and “stay” commands from its owner for ten minutes. A control group of dogs was simply left alone for ten minutes in cages. Then all the dogs were given a familiar toy with a sausage treat inside of it. All the dogs had played with the toy in the past and successfully extracted the treat, but for this study the toy was rigged so that the sausage could not be extracted. The control group of dogs spent several minutes trying to extract it, but the dogs who’d had to obey the commands gave up in less than a minute. It was the familiar ego-depletion effect, and the canine cure turned out to be familiar too. In a follow-up study, when the dogs were given different drinks, the drinks with sugar restored the willpower of the dogs who’d had to obey the commands. Newly fortified, they persisted with the toy just as long as the dogs who’d been in cages.

People with poor self-control were likelier to hit their partners and to commit a variety of other crimes, again and again, as demonstrated by June Tangney, who worked with Baumeister to develop the self-control scale on personality tests. When she tested prisoners and then tracked them for years after their release, she found that the ones with low self-control were most likely to commit more crimes and return to prison.

In one remarkable study, researchers in Finland went into a prison to measure the glucose tolerance of convicts who were about to be released. Then the scientists kept track of which ones went on to commit new crimes. Just by looking at the response to the glucose test, the researchers were able to predict with greater than 80% accuracy which convicts would go on to commit violent crimes. These men apparently had less self-control because of their impaired glucose tolerance, a condition in which the body has trouble converting food into usable energy.”

SO SOOOOO SKEPTICAL.

When people in laboratory experiments exercise mental self-control, their pulse becomes more erratic; conversely, people whose normal pulse is relatively variable seem to have more inner energy available for self-control, because they do better on laboratory tests of perseverance than people with steadier heartbeats.

Also pretty skeptical here, given past experience.

A psychoanalyst named Allen Wheelis in the late 1950s revealed what he considered a dirty little secret of his profession: Freudian therapies no longer worked the way they were supposed to. In his landmark book, The Quest For Identity, Wheelis described a change in character structure since Freud’s day. The Victorian middle-class citizens who formed the bulk of Freud’s patients had intensely strong wills, making it difficult for therapists to break through their ironclad defenses and their sense of what was right and wrong. Freud’s therapies had concenctrated on wayhs to break through and let them see why they were neurotic and miserable, because once those people achieved ionsight, they could change rather easily. By midcentury, thought, people’s character armor was different. Wheelis and his colleagues found that people achieved insight more qui9ckly than in Freud’s day, but then the therapy often stalled and failed. Lacking the sturdy character of the Victorians, people didn’t have the strength to follow up on the insight and change their lives

This is actually sorta plausible to me. My (very limited) experience with psychoanalysis is that it’s not nearly as hard as people claim to get patients to tell you things about themselves and produce apparent “revelations”, but these rarely change behavior in interesting ways. I’d never thought before that this might be a historical change as opposed to just Freud getting it wrong.

[Subjects] wore beepers that went off at random intervals seven times a day, prompting them to report whether they were currently experiencing some sort of desire…the researchers concluded that people spend at least a fifth of their waking hours resisting desires – between three and four hours a day…Overall, they succumbed to about a sixth of the temptations.

Someone responded by my last post by pointing out a Christian philosopher saying the same thing and noting that Science always thinks it’s gotten ahead of Religion only to find that Religion’s known it all along. Nevertheless, I will venture to say no religious source on temptation includes the observation that on average people succumb to about one-sixth of them. That one totally goes to Science.

# Book Review: The Machinery Of Freedom

[conflict of interest: David Friedman is an amazing person who has been very nice to me and among other things hosted the San Jose SSC meetup earlier this month]

David Friedman’s The Machinery of Freedom is half Libertarianism 101: Introduction To Libertarianism, and half Libertarianism 501: Technical Diagrams For Constructing An Anarcho-Capitalist State.

And aside from either of these, it’s interesting as a historical artifact. The first edition was published in 1973; the Third Edition copy I read is from last year, but the updates are minor and the book keeps its 1973 feel – including a discussion of health care economics which puts the price of a doctor’s visit at $10.

One of my takeaways was how new libertarianism was in 1973. The introduction says:

These peculiar views of mine are not peculiar to me. If they were, I would be paying Harper and Row to publish this book instead of Harper and Row paying me. My views are typical of the ideas of a small but growing group of people, a ‘movement’ that has begun to attract the attention of the national media. We call ourselves libertarians.

This book is concerned with libertarian ideas, not with a history of the libertarian movement or a description of its present condition. It is fashionable to measure the importance of ideas by the number and violence of their adherents. That is a fashion I shall not follow. If, when you finish this book, you have come to share many of my views, you will know the most important thing about the number of libertarians – that it is larger by one than when you started reading.

There is something very innocent about expecting someone to become a libertarian after reading a book arguing for libertarianism, something very much a product of the time when the movement was new and anything was possible. Friedman discusses and debates the views of Ayn Rand not as some sort of ascended cultural archetype, but as a fellow theorist who happens to be writing around the same time. It makes the book somehow fresher than one that starts from the perspective of “Okay, you’ve heard all of these arguments before, so let me preach to the choir and see what happens.”

But sometimes the book is dated in ways less innocuous than ten-dollar doctor visits. For example, in Chapter 5, “The Rich Get Richer And The Poor Get Richer,” Friedman argues against excessive concern with inequality, saying:

In absolute terms, the rich have gotten richer, but the gap between rich and poor seems, so far as very imperfect statistics make it possible to judge, to have been slowly closing…we can note that both the rise in the general standard of living and the decreasing inequality appear to have been occurring fairly steadily over a long period of time, in a variety of different more or less capitalist societies…in the previous chapter I argued that liberal measures tend to injure the poor, not benefit them, and to increase, not decrease inequality. If that has been true in the past, then the increasing equality we have experienced is in spite of, not because of, such measures…

Even if the capitalist invests all the income from his capital and consumes none of it, his wealth will only grow at the rate of return on capital. If the interest rate is less than the rate at which the total wages of workers increase, the relative wealth of the capitalists will decline. Historically, the rate of increase in total wages has run about 5 to 10 percent a year, roughly comparable to the interest rate earned by capital. Furthermore, capitalists consume part of their income; if they did not, there would be little point in being a capitalist. The share of the national income going to capital in this country has varied over time but not consistently increased, as shown in Appendix III.

The heartbreaking thing is that every word of this was true in 1973. In fact, 1973 is frequently given as the inflection point, when for some reason middle-class wages stopped rising at the same rate as the wealth of the top 1% and capital’s share of income started a steady climb (this is frequently blamed on Reagan, but started almost a decade before his presidency).

There are enough issues like this that they make the book’s arguments less compelling, or at least cry out to be addressed. Likewise, the book’s statistics are fascinating and in many cases very counterintuitive and convincing, but I have a lot of trouble double-checking them because they’re mostly 1970s statistics.

I can’t do justice to the Libertarian 101 arguments in this review because there are too many of them on too many different topics. This is too bad because they are excellent and fascinating and you should really read them. Aside from recommending you get the book, I’ll shove those into a separate Highlights post later this week. But for now I want to focus on the claim that I found most interesting: Government claims legitimacy partly from its role in helping the poor, but the costs fall disproportionately on the poor and it screws them over more than any other group:

Suppose that one hundred years ago someone tried to persuade me that democratic institutions could be used to transfer money from the bulk of the population to the poor. I could have made the following reply: “The poor, whom you wish to help, are many times outnumbered by the rest of the population, from whom you intend to take the money to help them. If the non-poor are not generous enough to give money to the poor voluntarily through private charity, what makes you think they will be such fools as to vote to force themselves to take it?”

I think I have a good answer to this question. Nobody’s vote makes very much difference, so people are happy to vote for signaling/psychological reasons rather than financial ones. If casting my vote to help the poor makes me feel like a good person, but losing money in redistribution schemes makes me poorer, well, my vote 100% determines whether I feel good or not, but only 1/300-million determines whether I get poorer. This might also be profitably mapped onto construal level theory, ie Robin Hanson’s Near Mode vs. Far Mode.

Anyway, having determined that democracy should not be expected to help the poor, he gets on to demonstrating that in fact it doesn’t:

There are some programs that give money to the poor – Aid to Families With Dependent Children, for instance. But such programs are vastly outweighed by those having the opposite effect – programs that injure the poor for the benefit of the not-poor. Almost surely, the poor would be better off if both the benefits that they now receive and the taxes, direct and indirect, that they now pay were abolished.

He then goes on to list examples, including Social Security, food subsidies (which increase food prices and go to rich farmers), state universities (since they cost tax money and mostly rich people go to university), and urban renewal projects (which bulldoze low-quality housing that the poor can afford to create high-quality housing that they can’t, thus pushing up their housing prices).

I don’t know much about the 1973 situation, but a lot of these don’t seem very convincing nowadays. Social Security no longer appears regressive: as per Wikipedia, “for people in the bottom fifth of the earnings distribution, the ratio of [Social Security] benefits to taxes is almost three times as high as it is for those in the top fifth.” And by my understanding, people who earn less than about $20,000 don’t pay federal income taxes at all, meaning the burden of universities, etc don’t fall upon them. A Cato Institute study finds that poor people on welfare can get benefits packages worth up to about $20,000. It seems really unlikely that whatever they have to pay because of farm subsidies or whatever compensates for that.

But Friedman also makes the stronger point that when government programs fail, it’s the poor who are most affected and who have the fewest other options. For example, he notes that the cost per capita of law enforcement/police/courts is $40 (remember, this is 1973!) and estimates that minus government waste and corruption, the free market could provide extremely competent policing for $20. He says:

There are many inhabitants of the ghetto who would be delighted to pay twenty dollars a year if in exchange they actually got protection; many of them have more than that stolen every year as a result of the poor protection they get from our government-run protection system. They would be even happier if at the same time they were relieved of the taxes that pay for the protection that the government police does not give them. In spite of popular myths about capitalism oppressing the poor, the poor are worst off in those things provided by government, such as schooling, police protection, and justice. There are more good cars in the ghetto than good schools.

I somewhat agree with the spirit of this quote, but certainly some of the problem is that poor people live in poor areas that collect little tax revenue and underfund their social services. Bigger government could solve this problem – just have school district funding set at the state or federal level. It’s less obvious that smaller government could – poor people would still have X dollars to spend on schools, for low values of X. But here we get into complicated proposals like vouchers and private policing that I’ll leave for later.

II.

Let’s get to what we’re really here for – the crazy anarcho-capitalist utopia.

This quote is very long, but it’s worth it:

How, without government, could we settle the disputes that are now settled in courts of law? How could we protect ourselves from criminals?

Consider first the easiest case, the resolution of disputes involving contracts between well-established firms. A large fraction of such disputes are now settled not by government courts but by private arbitration of the sort described in Chapter 18. The firms, when they draw up a contract, specify a procedure for arbitrating any dispute that may arise. Thus they avoid the expense and delay of the courts.

The arbitrator has no police force. His function is to render decisions, not to enforce them. Currently, arbitrated decisions are usually enforceable in the government courts, but that is a recent development; historically, enforcement came from a firm’s desire to maintain its reputation. After refusing to accept an arbitrator’s judgment, it is hard to persuade anyone else to sign a contract that specifies arbitration; no one wants to play a game of ‘heads you win, tails I lose’.

Arbitration arrangements are already widespread. As the courts continue to deteriorate, arbitration will continue to grow. But it only provides for the resolution of disputes over pre-existing contracts. Arbitration, by itself, provides no solution for the man whose car is dented by a careless driver, still less for the victim of theft; in both cases the plaintiff and defendant, having different interests and no prior agreement, are unlikely to find a mutually satisfactory arbitrator. Indeed, the defendant has no reason to accept any arbitration at all; he can only lose–which brings us to the problem of preventing coercion.

Protection from coercion is an economic good. It is presently sold in a variety of forms–Brinks guards, locks, burglar alarms. As the effectiveness of government police declines, these market substitutes for the police, like market substitutes for the courts, become more popular.

Suppose, then, that at some future time there are no government police, but instead private protection agencies. These agencies sell the service of protecting their clients against crime. Perhaps they also guarantee performance by insuring their clients against losses resulting from criminal acts.

How might such protection agencies protect? That would be an economic decision, depending on the’-costs and effectiveness of different alternatives. On the one extreme, they might limit themselves to passive defenses, installing elaborate locks and alarms. Or they might take no preventive action at all, but make great efforts to hunt down criminals guilty of crimes against their clients. They might maintain foot patrols or squad cars, like our present government police, or they might rely on electronic substitutes. In any case, they would be selling a service to their customers and would have a strong incentive to provide as high a quality of service as possible, at the lowest possible cost. It is reasonable to suppose that the quality of service would be higher and the cost lower than with the present governmental system.

Inevitably, conflicts would arise between one protective agency and another. How might they be resolved?

I come home one night and find my television set missing. I immediately call my protection agency, Tannahelp Inc., to report the theft. They send an agent. He checks the automatic camera which Tannahelp, as part of their service, installed in my living room and discovers a picture of one Joe Bock lugging the television set out the door. The Tannahelp agent contacts Joe, informs him that Tannahelp has reason to believe he is in possession of my television set, and suggests he return it, along with an extra ten dollars to pay for Tannahelp’s time and trouble in locating Joe. Joe replies that he has never seen my television set in his life and tells the Tannahelp agent to go to hell.

The agent points out that until Tannahelp is convinced there has been a mistake, he must proceed on the assumption that the television set is my property. Six Tannahelp employees, all large and energetic, will be at Joe’s door next morning to collect the set. Joe, in response, informs the agent that he also has a protection agency, Dawn Defense, and that his contract with them undoubtedly requires them to protect him if six goons try to break into his house and steal his television set.

The stage seems set for a nice little war between Tannahelp and Dawn Defense. It is precisely such a possibility that has led some libertarians who are not anarchists, most notably Ayn Rand, to reject the possibility of competing free-market protection agencies.

But wars are very expensive, and Tannahelp and Dawn Defense are both profit-making corporations, more interested in saving money than face. I think the rest of the story would be less violent than Miss Rand supposed.

The Tannahelp agent calls up his opposite number at Dawn Defense. ‘We’ve got a problem. . . .’ After explaining the situation, he points out that if Tannahelp sends six men and Dawn eight, there will be a fight. Someone might even get hurt. Whoever wins, by the time the conflict is over it will be expensive for both sides. They might even have to start paying their employees higher wages to make up for the risk. Then both firms will be forced to raise their rates. If they do, Murbard Ltd., an aggressive new firm which has been trying to get established in the area, will undercut their prices and steal their customers. There must be a better solution.

The man from Tannahelp suggests that the better solution is arbitration. They will take the dispute over my television set to a reputable local arbitration firm. If the arbitrator decides that Joe is innocent, Tannahelp agrees to pay Joe and Dawn Defense an indemnity to make up for their time and trouble. If he is found guilty, Dawn Defense will accept the verdict; since the television set is not Joe’s, they have no obligation to protect him when the men from Tannahelp come to seize it.

What I have described is a very makeshift arrangement. In practice, once anarcho-capitalist institutions were well established, protection agencies would anticipate such difficulties and arrange contracts in advance, before specific conflicts occurred, specifying the arbitrator who would settle them.

In such an anarchist society, who would make the laws? On what basis would the private arbitrator decide what acts were criminal and what their punishments should be? The answer is that systems of law would be produced for profit on the open market, just as books and bras are produced today. There could be competition among different brands of law, just as there is competition among different brands of cars.

In such a society there might be many courts and even many legal systems. Each pair of protection agencies agree in advance on which court they will use in case of conflict. Thus the laws under which a particular case is decided are determined implicitly by advance agreement between the protection agencies whose customers are involved. In principle, there could be a different court and a different set of laws for every pair of protection agencies. In practice, many agencies would probably find it convenient to patronize the same courts, and many courts might find it convenient to adopt identical, or nearly identical, systems of law in order to simplify matters for their customers.

Before labelling a society in which different people are under different laws chaotic and unjust, remember that in our society the law under which you are judged depends on the country, state, and even city in which you happen to be. Under the arrangements I am describing, it depends instead on your protective agency and the agency of the person you accuse of a crime or who accuses you of a crime.

In such a society law is produced on the market. A court supports itself by charging for the service of arbitrating disputes. Its success depends on its reputation for honesty, reliability, and promptness and on the desirability to potential customers of the particular set of laws it judges by. The immediate customers are protection agencies. But the protection agency is itself selling a product to its customers. Part of that product is the legal system, or systems, of the courts it patronizes and under which its customers will consequently be judged. Each protection agency will try to patronize those courts under whose legal system its customers would like to live.

The idea is that these protection agencies are companies like any other, and so will try to provide a good product at a low cost that satisfies their customers. People can choose their favorite, and so in some sense decide which laws to be bound by. Although they will not have complete flexibility in choosing their laws, lawmaking bodies will be sort of subject to consumer demand.

He correctly points out that contrary to what you might expect this system does not by definition exclude victimless crimes. If you want to hire a police agency that things being gay is a crime, you can pay them money to go find gay people and throw them out of town. Then the gay people will hire their own police agency to defend themselves. I think Friedman believes that opposing homosexuality has a major free rider problem, and that most people like to signal virtue by complaining about them but very few people would be willing to pay money for it. By comparison, gay people would be willing to pay a lot of money to be protected from this sort of thing, so their protection agencies would be stronger than the agencies of whoever wants to kick them out, and they’d stay.

This seems to me overly optimistic. After all, back when only a tiny percent of the country was tolerant of homosexuality, it might be that church groups could raise a lot of money to enforce anti-gay laws, and gay people were mostly poor and couldn’t raise very much money to defend themselves. I think I know what Friedman’s response would be, which is “Yes, and during that time in your real-world statist society, homosexuality was also illegal. Yes, you would have to wait for cultural norms to change before homosexuality would be legalized, but it would very likely be easier to do my way than yours.” I think he’s possibly right.

My overall conclusion is that I am delighted by this fascinating and elegant system and would very much like to see it tried somewhere very far away from me.

I am sure Friedman has to listen to so many objections that he can recite most of them by memory and is sick to death of them. Indeed, he admits this and devotes no small amount of space to rebutting many of them. Will we get taken over by one giant protection racket? Probably not, monopolies are rare in practice. Will criminals get their own protection and arbitration agencies that say crime is okay? Probably not; no other protection agency would agree to arbitrate on their terms, and without arbitration they would be in a war with all the other agencies, which the other agencies would win since legitimate business can mobilize more money than crime can. Would there be constant bloody battles? Probably not; profit-seeking corporations would be too smart to lose money that way when better options like arbitration are available. Would the heads of protection agencies form a pact, then use their combined might to take over the country and become kings? Probably not; right now police chiefs and military generals don’t do this, even though they are in a good position to.

Here are some objections of mine I didn’t see rebutted:

1. People who don’t purchase protection are pretty much fair game for anyone to rob or murder or torture or whatever. This seems harsh, especially since this society is likely to have a sizable underclass. I don’t know if “$20 for a year of police protection” was a reasonable estimate for the 70s, but I expect this would be much costlier now. Compare the percent of people who, pre-Obamacare, still didn’t have health insurance, and how much higher it would have been if there weren’t government programs that kind of got health insurance bundled in with employment.

2. Protection agencies are going to be engaged in constant brinksmanship for the same reason nation-states are engaged in constant brinksmanship. If Agency 1 wanted concessions from Agency 2, it has an incentive to seem kind of crazy and like it might actually declare real war, however unprofitable, in order to bluff Agency 2 into complying. Remember, countries have the same economic incentives to avoid war that companies do, but they still occasionally get involved in them. Even when they don’t, the threat of such leads many resources to be wasted in military buildup.

3. Security companies and their clients are very unlikely to want to pay for the cost of incarcerations. There’s no incentive to pay extra for criminal rights, so convicted criminals are likely to end up facing something like corporal punishment Never mind, this went an unexpected direction and is probably a good thing.

4. If I am the church-funded protection agency charged with flogging gay people, and you are the gay-person funded protection agency charged with protecting them, it’s hard to see what kind of arbitration we would agree on. I…uh…guess this might be another one that isn’t so bad, since that might mean the agencies are forced to actually fight, which raises the cost of being anti-gay to a potentially prohibitive level.

5. There are some things which might decrease crime in an area in general instead of just involving crime against a specific person. For example, adding streetlights, fighting drug abuse, putting troubled youth in after-school programs, fighting the broken window effect. If these are public goods, nobody will be incentivized to pay extra for them.

6. In fact, protection agencies have a strong incentive to make everybody as scared of crime as possible, and in fact to raise the actual crime rate if they can, in order to get people to buy their Premium plan. Given that this is anarcho-capitalism and there are no laws against crime, this can’t possibly end well.

7. It would be hard to have large-scale public laws. Right now Saudi Arabia can have laws about how no woman can go outside unveiled, America can have laws that nobody can go outside unclothed, and some European beaches can have laws saying go ahead and be naked. Likewise, some small villages can have zoning laws saying not to build non-scenic skyscrapers, but Dubai can say to build as high as you want and then some. This seems harder under anarcho-capitalism until people start coordinating the formation of intentional communities, at which point it becomes less anarcho-capitalism and more Patchwork.

8. Gang leaders and barbarian warlords had the chance to become protection agencies like this, but never did. This suggests that this system is unstable or unnatural. It’s possible that once the equilibrium of protection and arbitration agencies is established it will be stable, but of all of the various lawless societies to exist throughout history, none of them coalesced upon this system. Suspicious.

9. An extension of this: it’s unclear that we’re not already living in this society. It’s just that one protection and arbitration agency has completely taken over from all of the others and instituted a policy of using force against those who don’t pay for its services. That’s allowed under anarcho-capitalism because everything is allowed under anarcho-capitalism. So expecting anarcho-capitalism to be stable is expecting the thing that has already happened to not happen again a second time.

10. There seems to be a lot of opportunities for rich people to purchase greater privileges not available to the masses. After all, negotiation results are often determined by a party’s BATNA. Rich people may have access to very strong security companies (or premium plans from regular companies) that could win most fights; they can use this to insist on better arbitration terms. A rich person’s company might only accept basic arbitration (eg punish the rich person for murder) if other companies agree to lopsided deals (like don’t go after the rich person for less dramatic things like sexual harassment. On the other hand, a poorer person’s company might have to accept the worse side of the deal, where the poor person can be prosecuted for a very wide range of crimes against the rich person, including giving offense and not being respectful enough. Yes, it’s easy to see how a company could arise that charges extra in exchange for not accepting these compromises, but this still suggests you’re going to have more rights if you’re able to pay more money.

But the main reason I want this tried far away from me is none of these. It’s just a general expectation that something will go wrong when we try a social system we’ve never tried before. I was very impressed to learn that very few people predicted, before the fact, that Communist countries would have terrible economies. Even the American 1950s opponents of Communism argued that okay, fine, Communist countries will probably outperform capitalist countries economically, but freedom is more important than mere wealth.

If people can’t figure out that Communism might sink the economy, I don’t trust them to figure out all of the things that might go wrong with anarcho-capitalism. Even if David Friedman replies with utterly convincing rebuttals to all of my ten points above, it’s going to be the eleventh point I didn’t think of that makes the system explode.

III.

And this leads me into one of my deepest problems with libertarianism and anarcho-capitalism: why should it work?

I don’t mean the sort of “why should it work” where you answer with specific reasons why no, monopolies won’t form, and no, people won’t routinely sell themselves into slavery, and no, protection agencies won’t form a new feudal ruling class, and no, people won’t bash their heads against public goods problems and externalities forever without any market solutions appearing, and no, the poor won’t starve to death. I mean the very Outside View question of “why is it that, by coincidence, not using force is an effective way to solve all problems?”

Good governance is a really really hard problem. The idea that the solution to this problem contains zero bits of information, that it just solves itself if you leave people alone, seems astonishing. Even if we agree that capitalism works very well by incentivizing companies to do what the consumers wants, there are still a lot of peripheral issues which that just doesn’t cover. Friedman for example is a strong supporter of child rights, because children should mostly be free from coercion from their parents, and that children treated this way turn out better. Now in addition to solving governance with zero bits of information, you have solved optimal child-rearing with zero bits of information. That is implausibly impressive.

Given that the universe is allowed to throw whatever problems it wants at us, and that it has so far gleefully taken advantage of that right to come up with a whole host of very diverse and interesting ones, why is it that none of these problems are best addressed by a centralized entity with a monopoly on force? That seems like a pretty basic structure from a game-theoretic perspective, and you’re telling me it just never works in the real world? Shouldn’t there be at least one or two things where a government, or any form of coercive structure at all, is just the right answer? And can’t we just have a small government that does that?

The closest thing I’ve found to a response here is on page 142, where Friedman makes the following very witty observation:

The internal dynamic of limited government is something with which we, to our sorrow, have a good deal of practical experience. It took about 150 years, starting with a Bill of Rights that reserved to the states and the people all powers not explicitly delegated to the federal government, to produce a Supreme Court willing to rule that growing corn to feed your own hogs is interstate commerce and can therefore be regulated by Congress.

So if we have any kind of government at all, it will eventually metastasize into the sort of thing that makes laws about whether we’re allowed to grow corn to feed our own animals, or bans us from drinking raw milk, or whatever else it feels like doing.

So which is better: moving to full anarcho-capitalism, or trying to move towards a system that can provide more of the benefits of government with fewer of the costs?

I don’t know, so it’s a good thing I don’t have to choose. The obvious next step seems to be setting up anarcho-capitalist experiments somewhere and seeing how they do, as well as continuing to experiment with new and better forms of government. Trying to predict anything from theory runs into the same problem where everyone assumed Communism would be an economic powerhouse – we’re just not that smart. Instead we need to figure out ways to produce experimentation with and competition among different governments and government-like-entities – a goal I know David Friedman agrees with.

# List Of Passages I Highlighted In My Copy Of “Machinery Of Freedom”

Under any institutions, there are essentially only three ways that I can get another person to help me achieve my ends: love, trade, and force.

By love I mean making my end your end. Those who love me wish me to get what I want (except for those who think I am very stupid about what is good for me). So they voluntarily, ‘unselfishly’, help me. Love is too narrow a word. You might also share my end not because it is my end but because in a particular respect we perceive the good in the same way. You might volunteer to work on my political campaign, not because you love me, but because you think that it would be good if I were elected. Of course, we might share the common ends for entirely different reasons. I might think I was just what the country needed, and you, that I was just what the country deserved.

Love—more generally, the sharing of a common end—works well, but only for a limited range of problems. It is difficult to know very many people well enough to love them. Love can provide cooperation on complicated things among very small groups of people, such as families. It also works among large numbers of people for very simple ends—ends so simple that many different people can completely agree on them. But for a complicated end involving a large number of people—producing this book, for instance—love will not work. I cannot expect all the people whose cooperation I need—typesetters, editors, bookstore owners, loggers, pulpmill workers, and a thousand more—to know and love me well enough to want to publish this book for my sake. Nor can I expect them all to agree with my political views closely enough to view the publication of this book as an end in itself. Nor can I expect them all to be people who want to read the book and who therefore are willing to help produce it. I fall back on the second method: trade.

I contribute the time and effort to produce the manuscript. I get, in exchange, a chance to spread my views, a satisfying boost to my ego, and a little money. The people who want to read the book get the book. In exchange, they give money. The publishing firm and its employees, the editors, give the time, effort, and skill necessary to coordinate the rest of us; they get money and reputation. Loggers, printers, and the like give their effort and skill and get money in return. Thousands of people, perhaps millions, cooperate in a single task, each seeking his own ends. So under private property the first method, love, is used where it is workable. Where it is not, trade is used instead.

The attack on private property as selfish contrasts the second method with the first. It implies that the alternative to ‘selfish’ trade is ‘unselfish’ love. But, under private property, love already functions where it can. Nobody is prevented from doing something for free if he wants to. Many people—parents helping their children, volunteer workers in hospitals, scoutmasters—do just that. If, for those things that people are not willing to do for free, trade is replaced by anything, it must be by force. Instead of people being selfish and doing things because they want to, they will be unselfish and do them at the point of a gun.

Is this accusation unfair? The alternative offered by those who deplore selfishness is always government. It is selfish to do something for money, so the slums should be cleaned up by a ‘youth corps’ staffed via ‘universal service’. Translated, that means the job should be done by people who will be put in jail if they do not do it.

I just highlighted this because it was a beautifully phrased argument.

One of the most effective arguments against unregulated laissez faire has been that it invariably leads to monopoly. As George Orwell put it, “The trouble with competitions is that somebody wins them.” It is thus argued that government must intervene to prevent the formation of monopolies or, once formed, to control them. This is the usual justification for antitrust laws and such regulatory agencies as the Interstate Commerce Commission and the Civil Aeronautics Board.

The best historical refutation of this thesis is in two books by socialist historian Gabriel Kolko: The Triumph of Conservatism and Railroads and Regulation  
. He argues that at the end of the last century businessmen believed the future was with bigness, with conglomerates and cartels, but were wrong. The organizations they formed to control markets and reduce costs were almost invariably failures, returning lower profits than their smaller competitors, unable to fix prices, and controlling a steadily shrinking share of the market.

The regulatory commissions supposedly were formed to restrain monopolistic businessmen. Actually, Kolko argues, they were formed at the request of unsuccessful monopolists to prevent the competition which had frustrated their efforts.

So many books I need to read before I can have opinions on things.

It was in 1884 that railroad men in large numbers realized the advantages to them of federal control; it took 34 years to get the government to set their rates for them. The airline industry was born in a period more friendly to regulation. In 1938 the Civil Aeronautics Board (CAB), initially called the Civil Aeronautics Administration, was formed. It was given the power to regulate airline fares, to allocate routes among airlines, and to control the entry of new firms into the airline business. From that day until the deregulation of the industry in the late 1970s, no new trunk line— no major, scheduled, interstate passenger carrier—was started.

The CAB had one limitation: it could only regulate interstate airlines. There was one major intrastate route in the country— between San Francisco and Los Angeles. Pacific Southwest Airlines, which operated on that route, had no interstate operations and was therefore not subject to CAB rate fixing. Prior to deregulation, the fare between San Francisco and Los Angeles on PSA was about half that of any comparable interstate trip anywhere in the country. That gives us a good measure of the effect of the CAB on prices; it maintained them at about twice their competitive level.

In this complicated world it is rare that a political argument can be proved with evidence readily accessible to everyone, but until deregulation the airline industry provided one such case. If you did not believe that the effect of government regulation of transportation was to drive prices up, you could call any reliable travel agent and ask whether all interstate airline fares were the same, how PSA’s fare between San Francisco and Los Angeles compared with the fare charged by the major airlines, and how that fare compared with the fare on other major intercity routes of comparable length. If you do not believe that the ICC and the CAB are on the side of the industries they regulate, figure out why they set minimum as well as maximum fares.

Continuing to have nothing much to say except “wow”.

Defenders of [government health spending] programs argue that the poor are so poor they cannot afford vital medical care. Lurid reports to the contrary, most poor people are not on the edge of literal starvation; evidence indicates that in this country the number of calories consumed is virtually independent of income. If the poor spent more of their own money on doctors, they would not starve to death; they would merely eat worse, wear worse clothes, and live in even worse housing than they now do. If they do not spend very much money on medical care it is because that cost, which they are in an excellent position to evaluate, is too high.

Finally something where I can say something more interesting than wholehearted agreement.

The average cost of treatment for a heart attack is about $15,000. The poverty line for a single person in the US is $11,000. On the one hand, credit cards and loans can make up some of the difference; on the other, heart attacks are by no means even close to the most expensive medical condition. So if we’re talking about actually buying health care then no, the poor literally cannot afford it.

If we’re talking about buying health insurance, I understand a very cheap policy would cost about $2000, so the poor can probably literally afford that. I mean, they don’t have a whole lot of fat to trim, but they can afford it in the sense that if they choose to give up their home and car, and live on the streets, then they can have the health insurance. At least until their job fires them because they don’t have a car and can’t get there, and so they lose the money they were using to pay for it. But they won’t starve to death!

But then, why is starving to death such a uniquely interesting endpoint? Why assume that if the poor would die without health insurance we’re morally obligated to give it to them, but if they wouldn’t, we’re not? If we’re amoral or denying all obligations to help others, why care if the poor starve to death? And if we’re not amoral and feel some responsibility to the poor, why not also be concerned about them having a minimally tolerable life?

If some libertarian doesn’t think we have any obligation to help the poor, I’d rather they just say “Well, the poor might starve to death, but that’s too bad.”

Otherwise it seems sort of misleading to me. Saying “Well, the poor won’t literally starve to death” sounds like you’re saying “Well, it’s not that bad.” But if you were actually saying that, I could respond that it is that bad. It’s just bad in a non-starvation-related way. If you don’t care how bad it is, say so instead of hedging about whether starvation is occurring or not.

The best solution to this problem would be for any state instituting a voucher system to include, as part of the initial legislation, the provision that any institution can qualify as a school on the basis of the performance of its graduates on objective examinations. In New York, for instance, the law might state that any school would be recognized if the average performance of its graduating class on the Regents exam was higher than the performance of the graduating classes of the bottom third of the state’s public schools.

The best answer I’ve ever heard to the question of how to decide who gets school vouchers.

It might be possible to reform our present universities in the direction of such free-market universities. One way would be by the introduction of a ‘tuition diversion’ plan. This arrangement would allow students, while purchasing most of their education from the university, to arrange some courses taught by instructors of their own choice. A group of students would inform the university that they wished to take a course from an instructor from outside the university during the next year. The university would multiply the number of students by the average spent from each student’s tuition for the salary of one of his instructors for one quarter. The result would be the amount of their tuition the group wished to divert from paying an instructor of the university’s choice to paying an instructor of their own choice. The university would offer him that sum to teach the course or courses proposed. If he accepted, the students would be obligated to take the course.

The university would determine what credit, if any, was given for such courses. The number each student could take for credit might at first be severely limited. If the plan proved successful, it could be expanded until any such course could serve as an elective. Departments would still decide whether a given course would satisfy specific departmental requirements.

A tuition diversion plan does not appear to be a very revolutionary proposal; it can begin on a small scale as an educational experiment of the sort dear to the heart of every liberal educator. Such plans could, in time, revolutionize the universities.

At first, tuition diversion would be used to hire famous scholars on sabbatical leave, political figures of the left or right, film directors invited by college film groups, and other such notables. But it would also offer young academics an alternative to a normal career. Capable teachers would find that, by attracting many students, they could get a much larger salary than by working for a university. The large and growing pool of skilled ‘free-lance’ teachers would encourage more schools to adopt tuition diversion plans and thus simplify their own faculty recruitment problems. Universities would have to offer substantial incentives to keep their better teachers from being drawn off into free-lancing. Such incentives might take the form of effective market structures within the university, rewarding departments and professors for attracting students. Large universities would become radically decentralized, approximating free-market universities. Many courses would be taught by free-lancers, and the departments would develop independence verging on autarchy.

Under such institutions the students, although they might have the help of advisory services, would have to take the primary responsibility for the structure of their own education. Many students enter college unready for such responsibility. A competitive educational market would evolve other institutions to serve their needs. These would probably be small colleges offering a highly structured education with close personal contact for students who wished to begin their education by submitting to a plan of study designed by those who are already educated. A student could study at such a college until he felt ready to oversee his own education and then transfer to a university.

It is time to begin the subversion of the American system of higher schooling, with the objective not destruction but renaissance.

One of the better university reform proposals I’ve heard, plus an incremental strategy for achieving it!

I have solved the problem of urban mass transit. To apply rny solution to a major city requires a private company willing to invest a million dollars or so in hardware and a few million more in advertising and organization. The cost is low because my transit system is already over 99 percent built; its essence is the more efficient use of our present multibillion dollar investment in roads and automobiles. I call it jitney transit; it can most easily be thought of as something between taxicabs and hitch-hiking. Jitney stops, like present-day bus stops, would be arranged conveniently about the city. A commuter heading into town with an empty car would stop at the first jitney stop he came to and pick up any passengers going his way. He would proceed along his normal route, dropping off passengers when he passed their stops. Each passenger would pay a fee, according to an existing schedule listing the price between any pair of stops.

Holy !@#$, I think he has solved the problem of urban mass transit. There’s an obvious Uber parallel, but this system seems even better since it’s run by people going that direction anyway and each car will be packed, making the costs probably much cheaper. This is such an obviously good idea that I can only assume that it was regulation and the taxi lobby that prevented it from coming to pass. This paragraph probably did more to raise my confidence that there are extremely good libertarian solutions to important problems that we’re missing out on than anything else in the entire book.

Urban renewal uses the power of the government to prevent slums from spreading, a process sometimes referred to as ‘preventing urban blight’. For middle-class people on the border of low-income areas, this is valuable protection. But ‘urban blight’ is precisely the process by which more housing becomes available to low-income people. The supporters of urban renewal claim that they are improving the housing of the poor. In the Hyde Park area of Chicago, where I have lived much of my life, they tore down old, low-rental apartment houses and replaced them with $30,000 and $40,000 town houses. A great improvement, for those poor with $30,000. And this is the rule, not the exception, as was shown years ago by Martin Anderson in The Federal Bulldozer

I don’t know much about urban renewal programs or whether they purport to help the poor; anyone want to weigh in here?

Most conservatives now seem to have accepted, even embraced, the space program and with it the idea that the exploration of space can only be achieved by government. That idea is false. If we had not been in such a hurry, we not only could have landed a man on the moon, we could have done it at a profit.

How? Perhaps as a television spectacular. The moon landing alone had an audience of 400 million. If pay TV were legal, that huge audience could have been charged several billion dollars for the series of shows leading up to, including, and following the landing. If the average viewer watched, altogether, twenty hours of Apollo programs, that would be about 25 cents an hour for the greatest show off earth…

A greedy capitalist could have sold the moon landing in 1969 for something over $5 billion. The government spent $24 billion to get to the moon. It costs any government at least twice as much to do anything as it costs anyone else. It would have cost something under $12 billion to produce the Apollo program privately.

But Apollo was a crash program. If we had been in less of a hurry, it would have cost far less. While we were waiting, economic growth would increase the price for which the moon landing could be sold and technological progress would cut the cost of getting there. We would have arrived, at a profit, sometime in the seventies.

This is the business model of Mars One, which may be a scam. Which makes me wonder: how come, if the business model is sound, in 25 years of us having approximately the technology necessary to go to Mars, no one has come up with a non-scam version of this?

The NFL makes $10 billion a year through TV ads and sponsorship rights. I don’t know if a Mars mission would do better or worse than that – certainly the touchdown would be more exciting, but would people tune in month after month for “Yup, we’re still in this capsule, it’s really cramped in here and outside the window it just looks black”?

Robert Zubrin says he thinks a private company could reach Mars for $5 billion, which sounds promising, but he gets that because the government estimate is $50 billion and he thinks private companies can be ten times more efficient. Come on, Robert Zubrin! Even David Friedman estimates more like twice as efficient. I also note that SpaceX is estimating $1 billion to convert their existing Dragon to a crew-ready Dragon. $1 billion for a famously efficient private company to go from existing small rocket + small capsule to slightly improved small rocket + small capsule that can go to low Earth orbit – and you’re expecting another private company, right out of the gate, to be able to create ex nihilo a Mars-worthy spacecraft and the rocket that can launch it for $5 billion? Plus the astronaut training program, the production of the TV specials, the overhead for this new giant aerospace company you’re founding, the cost of the colony itself, etc, etc? Really?

And even if it’s possible in theory, think about the risk. The risk that the spacecraft explodes on the launch pad, and either you’ve just stuck your company name on a national tragedy or else you’d invested $6 billion in a TV special that’s never going to happen. Or the risk that five years later, the Mars One people come to you and say “Okay, Robert Zubrin was way too optimistic, we spent all your money to build the spacecraft’s left navigational fin, can you give us some more?” The risk that the Chinese beat you there and televising the second manned Mars landing isn’t very exciting.

Nothing I’ve seen so far convinces me that a serious version of Mars One is anywhere on the horizon. SpaceX will probably send a man to Mars someday, but they’ll do it because Elon Musk is vision-driven instead of profit-driven and he’s making enough profits somewhere else to fund his vision. And I don’t think even that would have worked without the funding and help that NASA has given SpaceX so far.

I worry that very big high-risk projects are exactly the sort of thing our current market system is really bad at.

My own conclusion—that drug companies should be free to sell, and their customers to buy, anything, subject to liability for damages caused by misrepresentation—must seem monstrous to many people. Certainly it means accepting the near certainty of a few people a year dying from unexpected side effects of new drugs.

This probably needs its own post, but no no no no no no no no, regulating drugs by liability is not a good idea, maybe even a worse idea than regulating them with regulations. Just as a quick example, here is an excerpt from Wikipedia’s article on the National Vaccine Injury Compensation Program:

“In 1988, the National Vaccine Injury Compensation Program (VICP) went into effect to compensate individuals and families of individuals who have been injured by covered childhood vaccines.[5] The VICP was adopted in response to an earlier scare over the pertussis portion of the DPT vaccine. These claims were later generally discredited, but some U.S. lawsuits against vaccine makers won substantial awards; most makers ceased production, and the last remaining major manufacturer threatened to do so. “

In other words, people kept winning so much money by suing the makers of pertussis vaccines that all of them except one just gave up and went out of business, and the only way the government saved that last one was by promising that the public purse would pay all of its losses. If the government hadn’t stepped in, we would not have vaccines right now because lawsuits would have made it unprofitable to make them. Idiotic lawsuits, I might add – pertussis vaccine doesn’t actually hurt people in any way. This is “my kid got autism after getting a vaccine” level stuff, and the courts were just like “Sure, fine, we believe you, let’s make the vaccine companies pay you so much money they all go bankrupt.”

This is not an isolated incident. The way malpractice works these days is that patients sue for things that are completely medically impossible, the malpractice insurances know that juries are too dumb to realize this, and they settle for more money than you will ever make honestly in your life. The FDA and its regulations are actually a rare force limiting this madness – if nothing else, a doctor can say “Well, that drug was approved by the FDA, so I wasn’t negligent in prescribing it to you.”

I understand that this book’s proposals include a large package of reforms which include those to the court system. But Friedman’s worries about how any “limited government” will eventually regrow into the kind of government that says you feeding your own grain to your own pigs is interstate commerce, are matched by my worries about how any “reformed court system” will eventually regrow into the kind of court system where children must be banned from sledding because if they get hurt they can sue the city for not having banned sledding, or lots of people who come to a psych hospital have to be committed lest years later somebody sue the hospital for not committing them.

If you invite more lawyers in to help control the government, you might end up like that Irish warlord who invited the English in to help control a rival warlord; you’ll find they’re even worse and they never leave.

The argument of this chapter received striking support in 1981, when the FDA published a press release confessing to mass murder. That was not, of course, the way in which the release was worded; it was simply an announcement that the FDA had approved the use of timolol, a ß-blocker, to prevent recurrences of heart attacks. At the time timolol was approved, ß-blockers had been widely used outside the U.S. for over ten years. It was estimated that the use of timolol would save from seven thousand to ten thousand lives a year in the U.S. So the FDA, by forbidding the use of ß-blockers before l981, was responsible for something close to a hundred thousand unnecessary deaths.

If examples of times when bad FDA decisions cost tens of thousands of lives made people abolish the FDA, we would probably have like negative seventeen FDAs by now.

Special interest politics is a simple game. A hundred people sit in a circle, each with his pocket full of pennies. A politician walks around the outside of the circle, taking a penny from each person. No one minds; who cares about a penny? When he has gotten all the way around the circle, the politician throws fifty cents down in front of one person, who is overjoyed at the unexpected windfall. The process is repeated, ending with a different person. After a hundred rounds everyone is a hundred cents poorer, fifty cents richer, and happy.

Annnnd we’re back to me just highlighting passages for rhetorical brilliance.

How much would it cost workers to purchase their firms? The total value of the shares of all stocks listed on the New York Stock Exchange in 1965 was $537 billion. The total wages and salaries of all private employees that year was $288.5 billion. State and federal income taxes totalled $75.2 billion. If the workers had chosen to live at the consumption standard of hippies, saving half their after-tax incomes, they could have gotten a majority share in every firm in two and a half years and bought the capitalists out, lock, stock, and barrel, in five. That is a substantial cost, but surely it is cheaper than organizing a revolution. Also less of a gamble. And, unlike a revolution, it does not have to be done all at once. The employees of one firm can buy it this decade, then use their profits to help fellow workers buy theirs later.

When you buy stock, you pay not only for the capital assets of the firm—buildings, machines, inventory, and the like —but also for its experience, reputation, and organization. If workers really can run firms better, these are unnecessary; all they need are the physical assets. Those assets—the net working capital of all corporations in the United States in 1965—totalled $171.7 billion. The workers could buy that much and go into business for themselves with 14 months’ worth of savings.

Compare to A Future For Socialism. In the research for that post I believe I found that the ratio of capital assets to wages had been rising pretty sharply recently, so it might take more time these days. But even if it took an entire decade, that’s a lot faster than most Communists expect the Revolution to come.

It probably says something very important about human nature and politics that the Socialist movement isn’t dominated by the project of doing exactly this.

# Is Everything A Religion?

I.

On the last Links thread, Eric Raymond claims that environmentalism is a religion. It has “sins” like wasting energy and driving gas-guzzling SUVs. It has “taboos” like genetically modified foods. It has an “apocalypse” in the form of global warming. It even has “rituals” in the form of weekly recycling.

This reminds me of an article I read recently claiming that transhumanism is a religion. But also of the article claiming that social justice is a religion. Also, liberalism is a religion. And conservativism is a religion. Libertarianism is a religion. Communism is a religion. Capitalism is like a religion. Objectivism is a religion. An anthropologist “confirms” that Apple is a religion. But UNIX is also a religion (apparently Linux was the Protestant Reformation).

Is there anything that isn’t like a religion? I spent this morning trying to come up with the least religious things I could think of. Trying to think of practical disciplines aimed at producing a quantifiable result, disciplines which strive to be evidence-based with a minimum of extraneous ideology. What came to mind was investing and medicine.

But investing is about propitiating a mysterious deity (the market) whose blessing or wrath bestows innumerable riches or total ruin. Believers follow gurus like Warren Buffett and Jim Cramer who promise that if they do the right things they will achieve financial salvation. Those who follow their pronouncements will enjoy the blissful afterlife of a comfortable retirement; those who violate their laws will spent their retirement in penury among much wailing and gnashing of teeth.

And medicine involves petitioners going to white-robed priests (doctors) who consult the holy scriptures (Harrison’s Clinical Medicine) to tell them how to live their lives. It has rituals (the yearly physical), taboos (smoking, overeating), and heretics (alternative medicine). Those who follow its rules are assured of a long, happy life; those who violate the rules of its priests will get cancer and die.

Maybe we’re still being too abstract here. What about, I don’t know, not stepping in front of buses? It certainly has a commandment (thou shalt not step in front of buses). It has notions of sin (stepping in front of buses) and virtue (not doing that). It has its rituals (looking both ways before you cross the street), its priests demanding obedience (crossing-guards), and its holy places (crosswalks). It promises blessings on the virtuous, but also terrible vengeance on the wicked (if you step in front of a bus, there will be much wailing and gnashing of teeth).

So one critique of these accusations is that “religion” is a broad enough category that anything can be mapped on to it:

Does it have well-known figures? Then they’re “gurus” and it’s a religion.

Are there books about it? Then those are “scriptures” and it’s a religion.

Does it recommend doing anything regularly? Then those are “rituals” and it’s a religion.

How about just doing anything at all? Then that’s a “commandment” and it’s a religion.

Does it say something is bad? Then that’s “sin” and it’s a religion.

Does it hope to improve the world, or worry about the world getting worse? That’s an “eschatology” and it’s a religion.

Do you disagree with it? Then since you’ve already determined all the evidence is against it, people must believe it on “faith” and it’s a religion.

II.

But that critique goes just a little too far. Once Communists start offering animal sacrifices to statues of Mao and requiring everyone own a copy of the Little Red Book and treat it respectfully, something is going on that’s deeper than just “it has well-known figures”.

Even though it’s easy to say that every belief or movement can be analogized to a religion, I still feel an intuition that some are more “religious” than others. Environmentalism and social justice seem more religious than gun control and pro-choice, even though all four are equally important lefty issues.

The first two are just more of a world-view. I can totally imagine someone saying “My life philosophy is centered around my passion for the environment”, but not so much “My life philosophy is centered around gun control.” I can see a speaker at a wedding saying “John and Jane are perfect for each other, since they are united by their shared passion about social justice”, but not so much “John and Jane are perfect for each other, since they are united by their shared passion for abortion rights.”

Both social justice and environmentalism spawn entire genres of art and literature, and I know people who pretty much exclusively draw their artistic consumption from those genres. But if somebody said “All of my art has a pro-choice theme”, that would probably be pretty creepy.

I know social justice people whose social circle is almost 100% based on social justice, and environmentalists whose social circle is almost 100% based on environmentalism. I don’t think there are that many people whose social circle is 100% based on gun control. And if someone says “I’m fanatical about the environment”, I get a whole lot of stereotypes about them – she probably eats granola, drives a Prius with a dreamcatcher in the window, has a college degree, does yoga. He probably goes hiking a lot, has a beard, takes supplements, is pretty relaxed. If someone says “I’m fanatical about gun control”, I’m stumped.

But all of this stuff about stereotypes and art and insularity sounds a little like religion but even more like culture, or at least subculture.

The difference between “religion” and “culture” has always been pretty vague. Shinto is the best example; it’s less a coherent metaphysical narrative than a bunch of things Japanese people do and a repository for Japanese traditions and rituals. A quick look at Hinduism reveals that they have no idea what gods they believe in, it’s a bunch of different religions stuck together under one umbrella, but the point is that it’s the sort of thing Indian people do and a repository of Indian traditions. Even though Jews have a pretty coherent religion, the line between “Jewish culture” and “Jewish religion” is equally fuzzy. Religion as distinct from culture seems like a pretty Western phenomenon, the result of a triumphant Christianity colonizing cultures it never originated from, ending out with the modern conception of culture as ethnic food + silly costumes.

American culture is paper-thin compared to say Hindu Indian culture, but consider its rituals like the Pledge of Allegiance, its holidays like the Fourth of July, its saints/culture heroes like George Washington and Benjamin Franklin, its myths like Paul Bunyan and Johnny Appleseed, its veneration of founding documents (the Declaration of Independence, the Constitution), and even its hymns like “America the Beautiful” and “Yankee Doodle”.

(the last of which, like all good hymns, uses such archaic language that almost nobody knows what the heck it means)

This gets called American civil religion a lot, but at this point I’m starting to wonder why it should. Maybe instead of accusing every culture of becoming a religion, we should just admit that our current concept of “religion” actually owes a lot to “culture”.

Eliezer writes that every cause wants to be a cult, but I’m not sure I agree with the connotations. I would say every cause wants to be a community. Communities hold values in common. Communities have rules their members have to follow. Communities have heroes and hierarchies. Communities shun people who don’t fit in.

And if all of this sounds super-conservative, keep in mind we’re still talking about environmentalism here, or social justice here. Values in common? Check. Rules? God yes. Heroes and hierarchies? You bet. Shunning people? All the time.

Communities and cultures have their share of danger. Their mix of social and epistemological functions means that any evidence challenging the community’s core beliefs will be taken as an attack on the members’ identity. As a result, community members risk ending up mind-killed. That’s not news. And I don’t think this is especially different from the way religious fanatics are mind-killed. And certainly someone could argue that “religion” is the perfect name for a culture built on shared belief.

But I still think it’s unfair to call these communities/cultures “religions”. “Religion” is too easy to use as the Worst Argument In The World here. It’s supposed to imply all of these other connotations of “religion” like “their beliefs are based on magical thinking” and “they use blind faith instead of reason” and “instead of coming up with a world-view based on evidence they just played Bible Mad Libs.” If those are the connotations you’ve got with “religion”, then I think the word “religion” is actively doing harm here, and you should just use “belief-based community” or “movement” or whatever.

# Extremism In Thought Experiment Is No Vice

[content warning: description of fictional rape and torture.]

Phil Robertson is being criticized for a thought experiment in which an atheist’s family is raped and murdered. On a talk show, he accused atheists of believing that there was no such thing as objective right or wrong, then continued:

I’ll make a bet with you. Two guys break into an atheist’s home. He has a little atheist wife and two little atheist daughters. Two guys break into his home and tie him up in a chair and gag him.

Then they take his two daughters in front of him and rape both of them and then shoot them, and they take his wife and then decapitate her head off in front of him, and then they can look at him and say, ‘Isn’t it great that I don’t have to worry about being judged? Isn’t it great that there’s nothing wrong with this? There’s no right or wrong, now, is it dude?’

Then you take a sharp knife and take his manhood and hold it in front of him and say, ‘Wouldn’t it be something if [there] was something wrong with this? But you’re the one who says there is no God, there’s no right, there’s no wrong, so we’re just having fun. We’re sick in the head, have a nice day.’

If it happened to them, they probably would say, ‘Something about this just ain’t right’.

The media has completely proportionally described this as Robinson “fantasizing about” raping atheists, and there are the usual calls for him to apologize/get fired/be beheaded.

So let me use whatever credibility I have as a guy with a philosophy degree to confirm that Phil Robertson is doing moral philosophy exactly right.

There’s a tradition at least as old as Kant of investigating philosophical dilemmas by appealing to our intuitions about extreme cases. Kant, remember, proposed that it was always wrong to lie. A contemporary of his, Benjamin Constant, made the following objection: suppose a murderer is at the door and wants to know where your friend is so he can murder her. If you say nothing, the murderer will get angry and kill you; if you tell the truth he will find and kill your friend; if you lie, he will go on a wild goose chase and give you time to call the police. Lying doesn’t sound so immoral now, does it?

The brilliance of Constant’s thought experiment lies in its extreme nature. If a person says they think lying is always wrong, we have two competing hypotheses: they’re accurately describing their own thought processes, which will indeed always output that lying is wrong; or they’re misjudging their own thought processes and actually there are some situations in which they will judge lying to be ethical. In order to distinguish between the two, we need to come up with a story that presents the strongest possible case for lying, so that even the tiniest shred of sympathy for lying can be dragged up to the surface.

So Constant says “It’s a murderer trying to kill your best friend”. And even this is suboptimal. It should be a mad scientist trying to kill everyone on Earth. Or an ancient demon, whose victory would doom everyone on Earth, man, woman, and child, to an eternity of the most terrible torture. If some people’s hidden algorithm is “lie when the stakes are high enough”, there we can be sure that the stakes are high enough to tease it out into the light of day.

Compare Churchill:

Churchill: Madam, would you sleep with me for five million pounds?  
Lady: Well, for five million pounds…well…that’s a lot of money.  
Churchill: Would you sleep with me for five pounds?  
Lady: (enraged) What kind of a woman do you think I am‽  
Churchill: We’ve already established what kind of a woman you are, now we’re just haggling over the price

The woman thinks she has a principle, “Never sleep with a man for money”. In fact, deep down, she believes it’s okay to sleep with a man for enough money. If Churchill had merely stuck to the five pounds question, she would have continued to believe she held the “never…” principle. By coming up with an extreme case (5 million Churchill-era pounds is about £250 million today) he was able to reveal that her apparent principle was actually a contingent effect of her real principle plus the situation.

In fact, compare physics. Physicists are always doing things like cooling stuff down to a millionth of a degree above absolute zero, or making clocks so precise they’ll be less than a second off by the time the sun goes out, or acclerating things to 99.99% of the speed of light. And one of the main reasons they do is to magnify small effects to the point where they can measure them. All movement is causing a little bit of time dilation, but if you want to detect it you need the world’s most accurate clock on the Space Shuttle when it’s traveling 25,000 miles per hour. In order to figure out how things really work, you need to turn things up to 11 so that the effect you want is impossible to miss. Everything in the universe has been exerting a gravitational effect on light all the time, but if you want to see it clearly you need to use the Sun during a solar eclipse, and if you really want to see it clearly your best bet is a black hole.

Great physicists and great philosophers share a certain perversity. The perversity is “Sure, this principle works in all remotely plausible real-world situations, but WHAT IF THERE’S A COMPLETELY RIDICULOUS SCENARIO WHERE IT DOESN’T HOLD??!?!” Newton’s theory of gravity explained everything from falling apples to the orbits of the planets impeccably for centuries, and then Einstein asked “Okay, but what if, when you get objects thousands of times larger than the Earth, there are tiny discrepancies in it, then we’d have to throw the whole thing out,” and instead of running him out of town on a rail scientists celebrated his genius. Likewise, moral philosophers are as happy as anyone else not to lie in the real world. But they wonder whether they might be revealed to be only simplifications of more fundamental principles, principles that can only be discovered by placing them in a cyclotron and accelerating them to 99.99% of the speed of light.

Sometimes this is even clearer than in the Kant example. Many people, if they think about it at all, believe that value aggregates linearly. That is, two murders are twice as much of a tragedy as one murder; a hundred people losing their homes is ten times as bad as ten people losing their homes.

Torture vs. Dust Specks is beautiful in its simplicity; it just takes this assumption and creates the most extreme case imaginable. Take a tiny harm and aggregate it an unimaginably high number of times; then compare it to against a big harm which is nowhere near the aggregated sum of the tiny ones. So which is worse, 3^^^3 (read: a number higher than you can imagine) people getting a single dust speck in their eye for a fraction of a second, or one person being tortured for fifty years?

Almost everybody thinks their principle is “things aggregate linearly”, but when you put it into relief like this, almost everybody’s intuition tells them the torture is worse. You can “bite the bullet” and admit that the dust specks are worse than the torture. Or you can throw out your previous principle saying that things aggregate linearly and try to find another principle about how to aggregate things (good luck).

Moral dilemmas are extreme and disgusting precisely because those are the only cases in which we can make our intuitions strong enough to be clearly detectable. If the question was just “Which is worse, a thousand people stubbing their toe or one person breaking their leg?” neither side would have been obviously worse than the other and our true intutition wouldn’t have come into sharp relief. So a good moral philosopher will always be talking about things like murder, torture, organ-stealing, Hitler, incest, drowning children, the death of four billion humans, et cetera.

Worse, a good moral philosopher should be constantly agreeing – or tempted to agree – to do horrible things in these cases. The whole point of these experiments is to collide two of your intuitions against each other and force you to violate at least one of them. In Kant’s example, either you’re lying, or you’re dooming your friend to die. In Jarvis’ Transplant Surgeon scenario, you’re either killing somebody to harvest their organs, or letting a whole hospital full of people die.

I once had someone call the torture vs. dust specks question “contrived moral dilemma porn” and say it proved that moral philosophers were kind of crappy people for even considering it. That bothered me. To look at moral philosophers and conclude “THESE PEOPLE LOVE TO TALK ABOUT INCEST AND ORGAN HARVESTING, AND BRAG ABOUT ALL THE CASES WHEN THEY’D BE OKAY DOING THAT STUFF. THEY ARE GROSS EDGELORDS AND PROBABLY FANTASIZE ABOUT HAVING SEX WITH THEIR SISTER ON THE HOSPITAL BED OF A PATIENT DYING OF END-STAGE KIDNEY DISEASE,” is to utterly miss the point.

So let’s talk about Phil Robertson.

Phil Robertson believes atheists are moral nihilists, or moral relativists, or something. He’s not quite right – there are a lot of atheists who are very moral realist – Objectivists, as their name implies, believe morality and everything else up to and including the best flavor of ice cream, is Objective – and even the atheists who aren’t quite moral realist usually hold some sort of compromise position where it’s meaningful to talk about right and wrong even if it’s not cosmically meaningful.

On the other hand – and I say this as the former secretary of a college atheist club who got to meet all sorts – there are a bunch of atheists who very much claim not to believe in morality. Less Wrong probably has fewer of them than the average atheist hangout, because we skew so heavily utilitarian, but our survey records 4% error theorists and 9% non-cognitivists. When Friendly Atheist says he “doesn’t know a single atheist or agnostic who thinks that terrorizing, raping, torturing, mutilating, and killing people is remotely OK”, I can believe that he doesn’t know one who would say so in those exact words. But I’m not sure how, for example, the error theorists could consistently argue against that position.

And what Phil Robertson does is exactly what I would do if I were debating an error theorist. I’d take the most gratuitiously horrible thing I could think of, describe it in the most graphic detail I could, and say “But don’t you think there’s something wrong with this?” If the error theorist says “no”, then I congratulate her for definitely being a real honest-to-goodness error theorist, and unless I can suddenly think up a way to bridge the is-ought dichotomy we’re finished. But if she says “Yes, it does seem like there should be something wrong there,” then we can start exploring what that means and whether error theory is the best framework in which to capture that intuition.

On the other hand, if I were debating Phil Robertson, I would ask him where he thinks morality comes from. And if he suggested some version of divine command theory, I could use an example of the graphic-horrifying-extreme-thought-experiment genre even older than Kant – namely, Abraham’s near-sacrifice of Isaac. If God commands you to kill your innocent child, is that the right thing to do? What if God commands you to rape and torture and mutilate your family? And it wouldn’t work if it were anything less extreme – if I just said “What if God told you to shoplift?” it would be easy to bite that bullet and he wouldn’t have to face the full implication of his views. But if I went with the extreme version? Maybe Robertson would find he’s not as big on divine command theory as he thought.

But this sort of discussion would only be possible if we could trust each other to take graphic thought experiments in the spirit in which they were conceived, and not as an opportunity to score cheap points.

[EDIT: This post was previously titled “High Energy Ethics”, but I changed it after realizing it was unintentionally lifted from elsewhere]

# Highlights From My Notes From Another Psychiatry Conference

I took a break from my busy schedule at Our Lady Of An Undisclosed Location to attend another local Psychiatry Conference.

This conference consisted of a series of talks about all the most important issues of the day, like ‘The Menace Of Psychologists Being Allowed To Prescribe Medication’, ‘How To Be An Advocate For Important Issues Affecting Your Patients Such As The Possibility That Psychologists Might Be Allowed To Prescribe Them Medication’, and ‘Protecting Members Of Disadvantaged Communities From Psychologists Prescribing Them Medication’.

As somebody who’s noticed that the average waiting list for a desperately ill person to see a psychiatrist is approaching the twelve month mark in some places, I was pretty okay with psychologists prescribing medication. The scare stories about how psychologists might prescribe medications unsafely didn’t have much effect on me, since I continue to believe that putting antidepressants in a vending machine would be a more safety-conscious system than what we have now (a vending machine would at least limit antidepressants to people who have $1.25 in change; the average primary care doctor is nowhere near that selective). Annnnnyway, this made me kind of uncomfortable at the conference and I Struck A Courageous Blow Against The Cartelization Of Medicine by sneaking out without putting my name on their mailing list.

But before I did, I managed to take some notes about what’s going on in the wider psychiatric world, including:

– The newest breakthrough in ensuring schizophrenic people take their medication (a hard problem!) is bundling the pills with an ingestable computer chip that transmits data from the patient’s stomach. It’s a bold plan, somewhat complicated by the fact that one of the most common symptoms of schizophrenia is the paranoid fear that somebody has implanted a chip in your body to monitor you. Can you imagine being a schizophrenic guy who has to explain to your new doctor that your old doctor put computer chips in your pills to monitor you? Yikes. If they go through with this, I hope they publish the results in the form of a sequel to The Three Christs of Ypsilanti.

– The same team is working on a smartphone app to detect schizophrenic relapses. The system uses GPS to monitor location, accelerometer to detect movements, and microphone to check tone of voice and speaking pattern, then throws it into a machine learning system that tries to differentiate psychotic from normal behavior (for example, psychotic people might speak faster, or rock back and forth a lot). Again, interesting idea. But again, one of the most common paranoid schizophrenic delusions is that their electronic devices are monitoring everything they do. If you make every one of a psychotic person’s delusions come true, such that they no longer have any beliefs that do not correspond to reality, does that technically mean you’ve cured them? I don’t know, but I’m glad we have people investigating this important issue.

– I’ll come out and say it: cluster randomization is really sketchy. Today I got to hear about a multi-center trial which randomized by location – half of their hospitals were the control group, the other half were the experimental group. Problem is, the patients in each hospital were given group-appropriate consent forms – either “We will be treating you as usual, but monitoring you more closely for a study” or “We will be giving you extra experimental treatment”. Not only does that break blinding, but it implies a different population of patients in each group – the ones willing to consent to monitoring versus the ones willing to consent to treatment? Might sicker people be more willing to sign the treatment consent, since they don’t want to deal with monitoring but treatment offers the chance for personal gain? Might paranoid people be more willing to sign the control consent, since they’re not being used as guinea pigs? I don’t know. But I checked those pre-intervention inter-group comparisons they have to show, and there were big differences between the two groups (for example, I think one – I can’t remember which – had like twice as many black people). Either randomize peopple properly or at least keep people blind to condition.

– On the other hand, I’m quickly losing my prejudice that RCTs always beat naturalistic studies. I’ll write more about this later, but today’s showcase was long-acting injectable versus oral antipsychotics. Conventional wisdom is that long-acting antipsychotics, in the right patient population, decrease relapse because they remove the option of not taking the medication. The best randomized controlled trials don’t find that. The best naturalistic epidemiological studies do. The expert who spoke today theorized – and I agree – that the naturalistic studies are right. He argued that one feature of RCTs is very close monitoring, which means the patients in them comply with their medication at an unnaturally high rate – thus removing the long-acting drugs’ one advantage. The studies conducted in the real world of patients not taking their medications regularly are more relevant.

– They say psychotic people don’t take their meds because they hate the side effects, or because they’re too crazy to know better, or because they just can’t be bothered. But one of the doctors today raised a novel hypothesis: are antipsychotics anti-addictive? After all, some of the most addictive drugs are those that raise dopamine levels – cocaine, meth, and MDMA are all either dopamine releasing agents or dopamine reuptake inhibitors. Antipsychotics have pretty much the opposite effect as those, lowering dopamine in the brain. Suspicious. But I have a feeling this isn’t true. Dopamine is more complicated than that. Levodopa-carbidopa, which is one step short of pure dopamine and is given to dopamine-deficient Parksinson’s patients, is as far as I know not addictive at all. It’s also very clearly antagonistic to antipsychotics. Probably antipsychotics are the opposite of non-addictive levodopa, not the opposite of cocaine or anything. I don’t know how to phrase it more rigorously than that. Still, I like the way that person thinks.

– Ever since Indiana’s legislature debated a bill that implied pi = 4, Midwestern states have had a reputation for trying to legislate science. Maybe this had something to do with the claim by one psychiatry lobbyist that Kansas’ legislature is trying to ban the DSM. I can’t find anything on it online and it sounds like an urban legend to me. Tangentially related silly clickbait: Arizona lawmakers say horses aren’t animals.

– Unintentional puns are some of my favorite puns. I still remember fondly when the head of a psychiatric hospital where I used to work said that if Obamacare passed there would be too many patients and the place would “turn into a madhouse”. I collected another good one today when an activist was talking about gun rights for psychiatric patients: “Taking guns from psychiatric patients isn’t going to be a panacea for violence – would anyone like to take a stab at why?”

– Clozapine really is the best antipsychotic, hands down, and the evidence isn’t even subtle. It’s also the most dangerous, and the rules say that you should only prescribe it to a patient after you’ve tried and failed with two other antipsychotics. One of the speakers was a researcher who’s trying to get a grant to prove that it’s actually more effective to try clozapine after only one failed antipsychotic, but the NIMH rejected his proposal because “even if you proved that, no one would listen”. They’re probably right. A lot of psychiatrists hate clozapine because it’s messy, scary, and requires a lot of paperwork and monitoring. The speaker presented survey after survey of psychiatrists making lame excuses like “My patients wouldn’t want it”, and then survey after survey of those psychiatrists’ patients saying they do so want it but nobody asked them. Clozapine is messy and scary and requires lots of paperwork, but if you’re a good doctor you’ll give your patient the drug that will help them anyway.

– The APA representative says that 95% of candidates supported by the APA’s PAC get elected. I think it was supposed to be a boast, like “look how effective we are”, but that’s a bit much. Either the APA single-handedly controls all American politics, or else they’re very careful to always back the winning side. Properly understood, that number should probably be taken as a measure of exactly how cynical they are.

– Not that they didn’t admit their cynicism straight out. Our Political Activism Consultant explained that state legislators are all sorta new and confused and inexperienced all the time because of term limits. And if you put on a nice suit and a tie and tell them “Hey, I’m a doctor from your district, here’s how you need to do health care policy…” you have a pretty good chance of getting them to nod along and assume you know what you’re doing. I didn’t realize how easy this was, and I hope I never use this power for evil.

– This is basically how the Eternal War Against Psychologists Being Allowed To Prescribe Medications is being fought, but the psychologists have caught on and now they have nice suits and ties too. Also, it turns out senators have a hard time differentiating the APA (American Psychiatric Association, fighting tooth and claw against psychologist prescribers) from the APA (American Psychological Association, fighting tooth and claw for psychologist prescribers) and they end up freaking out and trying to figure out why the same people are lobbying for both sides and whether this is some kind of weird shrink mind game thing.

– Drug companies were giving out stress brains! Like stress balls, only they’re shaped like brains and have little sulci and gyri on them! If in ten years I’m one of those people who never prescribes clozapine, it’ll because I’m prescribing the drug by the company that gave me a stress brain instead.

# Rational Orthography

What do DVORAK, polyamory, and home schooling have in common? They’re all about doing what’s weird-but-effective instead of what’s popular. What else is like that?

About three thousand years ago, the ancient Greeks invented a form of writing called boustrophedon. The first line was written left-to-right, the second right-to-left, and so on in a winding pattern. The advantage of the new system was that it was faster and easier to read – instead of constantly darting your eyes back and forth from one side of the page to the other at the end of each line, you just let them continue naturally.

The disadvantage was that it was hard to write, for much the same reasons most people would have trouble writing backwards now. So although boustrophedon and straight left-to-right Greek competed for a couple of centuries, in the end straight Greek won because the scribes were too lazy to do what was most convenient for their readers. According to Right Hand, Left Hand: The Origins of Asymmetry in Brains, Bodies, Atoms and Cultures:

A risk when offering any historical description is what has been called ‘the Whig interpretation of history’, the easy presumption that everything leads straightforwardly and inexorably to the highest state of humankind. Such an interpretation fails to look at the entire historical picture, ignoring the losers – in our case, the writing systems that became extinct […]

Boustrophedon, the writing of the ox, is, as it were, on the horns of a dilemma; either it is easier to read and more difficult to write, or vice versa. It is not surprising that it rapidly died out in ancient writing. Perhaps more surprising are moves to reintroduce it. Computers can be programmed so that only the standard twenty-six letters have to be typed on the keyboard, but the screen display or printout has normal or mirror-reversed letters according to the direction of the script. Enthusiasts claim boustrophedon is easier and quicker to read because the eye does not have to find its way back to the beginning of the next line.

These sorts of things have to start somewhere. So I asked SSC reader Bakkot to create a script that causes this blog to display in boustrophedon. I think you’ll agree that the experience is much improved. If there’s enough demand for a “classic” view, I can ask him to create some kind of optional browser add-on that will disable it, but I’d urge you to try the new version for a couple of weeks before turning to that “solution”.

More important: what if you want everything you read to be in boustrophedon from now on? For that I can unreservedly recommend The Boustrophedon Text Reader. Right now it only works on .txt, but hopefully as the movement catches on someone can turn it into a full browser extension.

# Chemical Imbalance

[content note: mental illness. I am still in training and do not understand these issues even as well as a fully-trained psychiatrist, let alone a researcher, so take all the biology and studies in here with a grain of salt until you double-check]

I.

IO9’s new article The Most Popular Antidepressants Are Based On An Outdated Theory jumps on a popular bandwagon of criticizing psychiatry for botching the “chemical imbalance” theory. See for example The New Yorker, BBC, The New York Times, and various books.

(…and also The Myth Of Chemical Imbalance, Debunking The Chemical Imbalance Myth, The Chemical Imbalance Fraud, and Depression Delusion, The Myth Of The Chemical Imbalance, etc)

According to all these sources psychiatry sold the public on antidepressants by claiming depression was just a chemical imbalance (usually fleshed out as “a simple deficiency of serotonin”) and so it was perfectly natural to take extra chemicals to correct it. However, they had no real evidence for this theory except that serotonergic drugs effectively treat depression, which is not very much evidence at all (antibiotics effectively treat pneumonia, but pneumonia isn’t “an antibiotic deficiency”). And now the research is unequivocal that serotonin deficiency is not the cause of depression, and psychiatry has ended up with lots of egg on its face.

This narrative is getting pushed especially hard by the antipsychiatry movement, who frame it as “proof” that psychiatrists are drug company shills who were deceiving the public. The conversation has required a host of rebuttals and counter-rebuttals.

For example here antipsychiatry blog Mad In America attemps to rebut psychiatrist Dr. Ronald Pies, who argues that psychiatrists never pushed the chemical imbalance theory. Pies says that “The ‘chemical imbalance theory’ was never a real theory, nor was it widely propounded by responsible practitioners in the field of psychiatry,” and cites the American Psychiatric Association’s 2005 statement on the causes of depression:

The exact causes of mental disorders are unknown, but an explosive growth of research has brought us closer to the answers. We can say that certain inherited dispositions interact with triggering environmental factors. Poverty and stress are well-known to be bad for your health—this is true for mental health and physical health. In fact, the distinction between “mental” illness and “physical” illness can be misleading. Like physical illnesses, mental disorders can have a biological nature. Many physical illnesses can also have a strong emotional component

Mad In America doesn’t accept his claim, and counter-cites two speeches by American Psychiatric Association presidents to prove that they did push the chemical imbalance theory:

In the last decade, neuroscience and psychiatric research has begun to unlock the brain’s secrets. We now know that mental illnesses – such as depression or schizophrenia – are not “moral weaknesses” or “imagined” but real diseases caused by abnormalities of brain structure and imbalances of chemicals in the brain.” – Richard Harding, 2001 APA president

And:

The way nerves talk to each other, and communicate, is through the secretion of a chemical called a neurotransmitter, which stimulates the circuit to be activated. And when this regulation of chemical neurotransmission is disturbed, you have the alterations in the functions that those brain areas are supposed to, to mediate. So in a condition like depression, or mania, which occurs in bipolar disorder, you have a disturbance in the neurochemistry in the part of the brain that regulates emotion. – Jeffrey Lieberman, 2012 APA President

I have no personal skin in this game. I’ve only been a psychiatrist for two years, which means I started well after the term “chemical imbalance” fell out of fashion. I get to use the excuse favored by young children everywhere: “It was like this when I got here”. But I still feel like the accusations in this case are unfair, and I would like to defend my profession.

I propose that the term “chemical imbalance” hides a sort of bait-and-switch going on between the following two statements:

(A): Depression is complicated, but it seems to involve disruptions to the levels of brain chemicals in some important way

(B): We understand depression perfectly now, it’s just a deficiency of serotonin.

If you equivocate between them, you can prove that psychiatrists were saying (A), and you can prove that (B) is false and stupid, and then it’s sort of like psychiatrists were saying something false and stupid!

But it isn’t too hard to prove that psychiatrists, when they talked about “chemical imbalance”, meant something more like (A). I mean, look at the quotes above by which Mad In America tries to prove psychiatrists guilty of pushing chemical imbalance. Both sound more like (A) than (B). Neither mentions serotonin by name. Both talk about the chemical aspect as part of a larger picture: Harding in the context of abnormalities in brain structure, Lieberman in the context of some external force disrupting neurotransmission. Neither uses the word “serotonin” or “deficiency”. If the antipsychiatry community had quotes of APA officials saying it’s all serotonin deficiency, don’t you think they would have used them?

Further, anyone who said that depression was caused solely by serotonin deficiency wouldn’t just be failing as a scientist, but also failing as a drug company shill. Pfizer spent billions of dollars on Effexor, which hits norepinephrine as well as serotonin, and they’re just going to dismiss all of that as useless? GlaxoSmithKline has Wellbutrin, which hits dopamine and norepinephrine and maybe acetylcholine but doesn’t get serotonin at all. So everyone, including the shills, especially the shills, has been very careful to say that depression was a “chemical imbalance” rather than a serotonin deficiency per se.

So if you want to prove that psychiatrists were deluded or deceitful, you’re going to have to disprove not just statement (B) – which never represented a good scientific or clinical consensus – but statement (A). And that’s going to be hard, because as far as I can tell statement (A) still looks pretty plausible.

II.

If you listen to these articles, psychiatrists decided that neurotransmitters (or just serotonin?) were implicated in depression solely on the evidence that SSRIs were effective antidepressants, even though every study trying to measure serotonin levels directly came back with negative results. For example, The Myth Of The Chemical Imbalance Theory writes:

There is no question that the chemical imbalance theory has spurred chemists to invent new anti-depressants, or that these anti-depressants have been shown to work; but proof that low serotonin is to blame for depression – and that boosting serotonin levels is the key to its treatment – has eluded researchers.

For starters, it is impossible to directly measure brain serotonin levels in humans. You can’t sample human brain tissue without also destroying it. A crude work-around involves measuring levels of a serotonin metabolite, 5-HIAA, in cerebrospinal fluid (CSF), which can only be obtained with a spinal tap. A handful of studies from the 1980s found slightly decreased 5-HIAA in the CSF of depressed and suicidal patients, while later studies have produced conflicting results on whether SSRIs lower or raise CSF levels of 5-HIAA. These studies are all circumstantial with regards to actual serotonin levels, though, and the fact remains there is no direct evidence of a chemical imbalance underlying depression.

The corollary to the chemical imbalance theory, which implies that raising brain serotonin levels alleviates depression, has also been hard to prove. As mentioned previously, the serotonin-depleting drug reserpine was itself shown to be an effective anti-depressant in the 1950s, the same decade in which other studies claimed that reserpine caused depression-like symptoms. At the time, few psychiatrists acknowledged these conflicting reports, as the studies muddled a beautiful, though incorrect, theory. Tianeptine is another drug that decreases serotonin levels while also serving as a bona-fide anti-depressant. Tianeptine does just the opposite of SSRIs – it enhances serotonin reuptake. Wellbutrin is a third anti-depressant that doesn’t increase serotonin levels. You get the picture.

If you prefer your data to be derived more accurately, but less relevantly, from rodents, you might consider a recent meta-analysis carried out by researchers led by McMaster University psychologist Paul Andrews. Their investigation revealed that, in rodents, depression was usually associated with elevated serotonin levels. Andrews argues that depression is therefore a disorder of too much serotonin, but the ambiguous truth is that different experiments have shown “activation or blockage of certain serotonin receptors [to improve] or worsen depression symptoms in an unpredictable manner.”

Other problems with the chemical imbalance model of depression have been well documented elsewhere. For instance, if low serotonin levels were responsible for symptoms of depression, it stands to reason that boosting levels of serotonin should alleviate symptoms more or less immediately. In fact, antidepressants can take more than a month to take effect. Clearly, something here just doesn’t add up.

Clearly!

GABA is a neurotransmitter that promotes inhibition and relaxation. Suppose I were to tell you that alcohol is a drug that mimics the effects of GABA. Which it is.

You might say: something is wrong with this theory! After all, people who drink alcohol don’t always get relaxed and inhibited. A lot of the time they get uninhibited and angry and violent! And then if they drink too much of it, they get super-inhibited to the point where they’re in a total blackout. Also, alcoholics who have been drinking for many years have higher levels of anxiety than non-alcoholics, but anxiety is also the opposite of relaxation! Clearly, something here just doesn’t add up. Maybe the neuroscientists are all shills for Budweiser!

Or else maybe the brain is kind of complicated. In the case of alcohol we pretty much know what’s going on. Alcohol does inhibit and relax you, but in some people and at some doses, it preferentially inhibits and relaxes the parts of the brain involved in inhibiting and relaxing the rest of the brain, meaning that the person as a whole because more uninhibited and violent. At higher doses, it inhibits and relaxes the entire brain, leading to confusion and eventually blackout. And once you’ve been taking alcohol for many years, your brain adjusts to the higher level of GABA-like chemicals by producing fewer GABA receptors, making you more anxious.in general. It’s a whole bunch of contradictory effects, but when you look at the neuroscience it makes sense.

We know less about the serotonin picture, but what we know suggests something similar is going on. Serotonin has different effects in lots of different parts of the brain. There are fourteen different types of serotonin receptor, all of which do subtly different things. Some serotonergic neurons have autoreceptors that cause decreased release of serotonin in response to serotonin. The brain responds to different levels of serotonin by slowly altering endogenous serotonin production as well as the expression of the different serotonin receptors. Etc, etc, etc.

Lest it sound like I’m making excuses rather than presenting evidence: A study on a monkey model – generally preferred to humans when you want to kill your patients and take apart their brains when you’re done – showed that depressed macaques had elevated levels of serotonin in the dorsal raphe nuclei and decreased levels of serotonin in the hippocampus, resulting in average levels of serotonin in the cerebrospinal fluid where the experiments mentioned above took their serotonin measurements. A study with a more sophisticated measurement process, Elevated Brain Serotonin Turnover in Patients With Depression, found that depressed subjects had serotonin turnover as measured in the jugular vein about twice as high as healthy controls (p = 0.003), and successful treatment with SSRI therapy corrected this imbalance (though others dispute the methodology).

All of this sort of fits. If depression involves a distorted pattern of serotonin across the brain, then both certain drugs that increase serotonin levels and certain drugs that decrease it might be helpful. And SSRIs might take a month to work if their mechanism of action isn’t the direct serotonin increase, but a contrary response they provoke from the brain. I think I heard from someone in the field that a month is about how long it takes for them to change the levels of expressed 5HT receptors by altering genetic transcription. Or something. I’m not a neuroscientist (though you can read some more complicated work from people who are) and I don’t know. The point is that you can get a heck of a lot more complex than just “Too little serotonin!” versus “Too much serotonin!”

So does this mean depression “was really serotonin after all”?

No. It means we have good evidence serotonin is involved somewhere. Among the other things that we have good evidence are involved somewhere are: dopamine, norepinephrine, acetylcholine, cytokines, BDNF, thyroid hormones, and whether the kids at school picked on you in first grade.

Suppose you ask me what caused you to become blind. I happen to have your medical records and know that the answer is proliferative retinopathy secondary to Type 2 diabetes, but you’ve been living in a cave your entire life and never even heard of diabetes. Which is the correct answer to your question?

1. Your blindness is caused by tiny little blood vessels growing all over your eyes  
2. Your blindness is caused by imbalance in a chemical called protein kinase C-delta and the resulting signaling cascade  
3. Your blindness is caused by too much sugar in your blood  
4. Your blindness is caused by your cells becoming less sensitive to insulin  
5. Your blindness is caused by you drinking too much Coca-Cola

All of these are true. You drink too much Coca-Cola, it causes your cells to lose insulin sensitivity, that causes too much sugar in the blood, that increases the activity of PKC-delta, and that causes little blood vessels to grow all over your eyes. Sometimes the chain is different. Maybe you drank too much lemonade instead of too much Coca-Cola. Maybe you drank too much Coca-Cola, but actually instead of causing diabetes it caused hypertension and then you got hypertensive retinopathy which made you blind. Maybe it was diabetic retinopathy, but actually you haven’t gotten to the proliferative stage yet, and you just had a lot of your blood vessels get damaged and start leaking and causing macular oedema. Maybe it was diabetic retinopathy, but you had a perfect diet and lost the genetic lottery. I don’t know.

If someone told you “We think it involves an imbalance in protein kinase” it would be woefully incomplete. But if someone said “That doctor there said your blindness was caused by an imbalance in protein kinase, that proves he’s a fraud!”, well, no, it wouldn’t.

Except the situation is even more complicated than this, because at least I specified this guy had diabetic retinopathy. What if somebody just asked “What causes blindness?” “High protein kinase” or “high blood sugar” would be two answers, and you could find tests supporting both. But “cataracts” would be another good answer. So would “people getting acid thrown in their eyes”.

All I’m saying is that depression is complicated. Discovering its relationship to the serotonin system is a lot like saying “blindness quite often has something to do with the retina”. It’s a big step forward, and don’t believe anyone who says it isn’t, but it’s not anywhere near the whole picture.

III.

And this starts to get into the next important point I want to bring up, which is chemical imbalance is a really broad idea.

Like, some of these articles seem to want to contrast the “discredited” chemical imbalance theory with up-and-coming “more sophisticated” theories based on hippocampal neurogenesis and neuroinflammation. Well, I have bad news for you. Hippocampal neurogenesis is heavily regulated by brain-derived neutrophic factor, a chemical. Neuroinflammation is mediated by cytokines. Which are also chemicals. Do you think depression is caused by stress? The stress hormone cortisol is…a chemical. Do you think it’s entirely genetic? Genes code for proteins – chemicals again. Do you think it’s caused by poor diet? What exactly do you think food is made of?

Diabetes is caused by a chemical imbalance: too much sugar (or too little insulin) in the blood. Parkinson’s is caused by a chemical imbalance: too little dopamine in the basal ganglia. Heart attacks are caused by a chemical imbalance: too many of the wrong kinds of lipids and lipid-related plaques in the coronary arteries.

I can get even more nitpicky if you want. The Donner Party died of chemical imbalance – too few fatty acids, proteins, and carbohydrates. The passengers of the Titanic died of a chemical imbalance – H2O in the lungs instead of O2. And it was a chemical imbalance that got Hiroshima in the end: excess uranium-235. Anything that’s not caused by ghosts is going to be “a chemical imbalance” in some sense of the word.

This is why I’m being so insistent that psychiatrists referred to “a chemical imbalance” rather than “a serotonin deficiency”. They were hedging the heck out of their bets. It might be BDNF, or cytokines, or whatever. But if something happens in the body and doesn’t show up as a gross anatomical defect on MRI, it’s a pretty good bet it’s chemical in some sense of the word.

So is this a giant cop-out? Psychiatrists said “it’s a chemical imbalance” to make it sound like they knew what they were talking about, when in fact all they meant was “it’s a thing that exists”?

Sort of.

Anything that isn’t caused by ghosts is going to be “a chemical imbalance” in some sense of the word. But in the latter half of the twentieth century, “depression is not caused by ghosts” was a revolutionary statement, and one that desperately needed to be said.

I still see this. People come in with depression, and they think it means they’re lazy, or they don’t have enough willpower, or they’re bad people. Or else they don’t think it, but their families do: why can’t she just pull herself up with her own bootstraps, make a bit of an effort? Or: we were good parents, we did everything right, why is he still doing this? Doesn’t he love us?

And I could say: “Well, it’s complicated, but basically in people who are genetically predisposed, some sort of precipitating factor, which can be anything from a disruption in circadian rhythm to a stressful event that increases levels of cortisol to anything that activates the immune system into a pro-inflammatory mode, is going to trigger a bunch of different changes along metabolic pathways that shifts all of them into a different attractor state. This can involve the release of cytokines which cause neuroinflammation which shifts the balance between kynurinins and serotonin in the tryptophan pathway, or a decrease in secretion of brain-derived neutrotrophic factor which inhibits hippocampal neurogenesis, and for some reason all of this also seems to elevate serotonin in the raphe nuclei but decrease it in the hippocampus, and probably other monoamines like dopamine and norepinephrine are involved as well, and of course we can’t forget the hypothalamopituitaryadrenocortical axis, although for all I know this is all total bunk and the real culprit is some other system that has downstream effects on all of these or just…”

Or I could say: “Fuck you, it’s a chemical imbalance.”

Last time I talked about the definition of disease I said that people want diseases to “be caused by the sorts of thing you study in biology: proteins, bacteria, ions, viruses, genes.”

I don’t think I could actually get away with telling a patient’s family “it’s caused by, you know, biology stuff” without them asking if I really went to medical school. I don’t think I’d use the term “chemical imbalance” precisely; too likely to trigger a knee-jerk reaction from people reading exactly these articles I’m responding to. But I think I would say something alone those lines. “We don’t know exactly, but it probably involves problems with brain structure and brain chemicals,” maybe. That covers about the same ground as “biology stuff” while also sounding like I’m at least trying to answer their question.

So if what I’m actually saying with that is “depression is caused by complicated biology stuff you don’t understand, and not by things like your son not really loving you, or being lazy,” am I sure that’s right?

I won’t say all depression is 100% caused by internal failures of biology in the same way that for example cystic fibrosis is caused 100% by internal failures of biology. I am happy to admit that some depressions can be caused by being in a crappy social situation, being abused as a child, being stuck in an unhappy marriage, being worried about problems at work, stuff like that.

But it’s far from obvious that being stuck in an unhappy marriage should drain your energy, drain your concentration, make you stop enjoying your hobbies, and finally drive you to suicide. We can imagine another person, or another way of designing a person, where someone says “I hate my husband, so I try to stay away from him as much as I can by working extra hard and spending my free time playing frisbee with my dog in the park.” But instead, someone hates their husband, and it drives all the joy out of their life to the point where they can’t go to work, they can’t play with their dog, they just sit around wishing they were dead.

And is that the fault of “biology stuff”? That’s a harder question than it sounds. What would it mean to say ‘no’? If we are strict materialists who don’t believe in some kind of division of labor between the brain and the soul, then yes, if it’s a feeling you’re having, it’s based in biology.

I’ve previously said we use talk of disease and biology to distinguish between things we can expect to respond to rational choice and social incentives and things that don’t. If I’m lying in bed because I’m sleepy, then yelling at me to get up will solve the problem, so we call sleepiness a natural state. If I’m lying in bed because I’m paralyzed, then yelling at me to get up won’t change anything, so we call paralysis a disease state. Talk of biology tells people to shut off their normal intuitive ways of modeling the world. Intuitively, if my son is refusing to go to work, it means I didn’t raise him very well and he doesn’t love me enough to help support the family. If I say “depression is a chemical imbalance”, well, that means that the problem is some sort of complicated science thing and I should stop using my “mirror neurons” and my social skills module to figure out where I went wrong or where he went wrong.

In other words, everything we do is caused by brain chemicals, but usually we think about them on the human terms, like “He went to the diner because he was hungry” and not “He went to the diner because the level of dopamine in the appetite center of his hypothalamus reached a critical level which caused it to fire messages at the complex planning center which told his motor cortex to move his legs to…” – even though both are correct. Very occasionally, some things happen that we can’t think about on the human terms, like a seizure – we can’t explain in terms of desires or emotions or goals an epileptic person is flailing their limbs, so we have to go down to the lower-level brain chemical explanation.

What “chemical imbalance” does for depression is try to force it down to this lower level, tell people to stop trying to use rational and emotional explanations for why their friend or family member is acting this way. It’s not a claim that nothing caused the chemical imbalance – maybe a recent breakup did – but if you try to use your normal social intuitions to determine why your friend or family member is behaving the way they are after the breakup, you’re going to get screwy results.

(in much the same way, if I just saw you take a giant handful of amphetamines, I pretty much know why you’re having a seizure, but I still can’t rationally / intuitively model the experience of why you’re “choosing” to move your limbs the way that you are.)

(though it’s important for me to temper this by mentioning that many people diagnosed with depression don’t have it)

There’s still one more question, which is: are you sure that depression patients’ experience is so incommensurable with healthy people’s experiences that it’s better to model their behavior as based on mysterious brain chemicals rather than on rational choice?

And part of what I’m going on is the stated experience of depressed people themselves. As for the rest, I can only plead consistency. I think people’s political opinions are highly genetically loaded and appear to be related to the structure of the insula and amygdala. I think large-scale variations in crime rate are mostly attributable to environmental levels of lead and probably other chemicals. It would be really weird if depression were the one area where we could always count on the inside view not to lead us astray.

So this is my answer to the accusation that psychiatry erred in promoting the idea of a “chemical imbalance”. The idea that depression is a drop-dead simple serotonin deficiency was never taken seriously by mainstream psychiatry. The idea that depression was a complicated pattern of derangement in several different brain chemicals that may well be interacting with or downstream from other causes has always been taken seriously, and continues to be pretty plausible. Whatever depression is, it’s very likely it will involve chemicals in some way, and it’s useful to emphasize that fact in order to convince people to take depression seriously as something that is beyond the intuitively-modeled “free will” of the people suffering it. “Chemical imbalance” is probably no longer the best phrase for that because of the baggage it’s taken on, but the best phrase will probably be one that captures a lot of the same idea.

# No Physical Substrate, No Problem

I.

Yesterday I posted a link to an article in which Steve Wozniak joins other luminaries like Elon Musk and Bill Gates in warning about the dangers of artificial superintelligence. A commenter replied:

Elon Musk, Stephen Hawking, Bill Gates, and Steve Wozniak still aren’t enough for me, not until one of them can describe the process by which we go from ‘AI exists on computer’ to ‘AI killing human beings in physical reality’ by using something other than ridiculous, unforgivable cheating.

There are lots of good arguments against considering superintelligence a threat. Maybe strong AI is centuries or millennia away. Maybe there will be a very gradual transition from human-level AI to superintelligent AI that no single agent will be able to exploit. And maybe superintelligence can be safely contained in a very carefully shielded chamber with no means of connection to the outside world.

But the argument above has always seemed to me like one of the weakest. Maybe we’ll create a superintelligence, but it will just have no idea how to affect the physical world, and will just have to stay forever trapped in a machine connected to a worldwide network of computers that control every aspect of our economic and social lives? Really?

Normal, non-superintelligent people have already used the Internet to make money, form mass movements, and hire others to complete tasks for them. We can assume a true superintelligence – a mind much smarter than we are – will be able to do all these things as well or better than any human.

II.

Satoshi Nakamoto already made a billion dollars online without anybody knowing his true identity just by being good at math and having a bit of foresight. He’s probably not an AI, but he could have been.

That’s assuming our hypothetical superintelligence doesn’t just hack into a couple big banks and transfer their money to itself – again something some humans have already made a billion dollars doing. And that’s assuming it doesn’t just invent a really useful program and then offer it as shareware – another tried-and-true way of becoming a billionaire. And even that’s assuming it doesn’t just get a reasonable amount of money, then invest it very cleverly – another thing humans have already become billionaires doing.

III.

Mohammed was never a billionaire, but he does have 1.57 billion followers (a superintelligence presumably wouldn’t repeat his mistake of dying before his movement really came into its own). The Prophet started at the bottom – converting his friends and family to Islam one by one – and grew exponentially from there. Although he had the unfair advantage of a physical body, there’s no reason he needed it – if he’d lived today, maybe he would have converted Ali over GChat or Skype. In any case, the poetry of the Koran and the zeal of his followers attracted far more people than his personal appearance ever could have.

Other gurus and religious leaders’ fame is even more transparently a result of their writing rather than their visible personality; consider Ayn Rand’s success in founding a powerful Objectivist movement out of the people who read her books. In fact, some of the most famous religious movements in history, from the Nation of Islam to Christianity itself, have been founded secondhand by disciples who relayed the words of a leader whose very existence is difficult to confirm.

What kind of a movement might be founded by a superintelligence with more spiritual creativity than Mohammed, better writing skills than Rand, the entire Internet to evangelize, and billions of dollars to spend spreading its message? The Church of Scientology is already powerful enough to intimidate national governments; imagine a vastly superior version founded not by a second-rate sci-fi writer but by an entity straight out of science fiction itself.

IV.

And really all of this talk of gathering money and power is kind of redundant. Far easier to just borrow somebody else’s.

Imagine an AI that emails Kim Jong-un. It gives him a carrot – say, a billion dollars and all South Korean military codes – and a stick – it has hacked all his accounts and knows all his most blackmail-able secrets. All it wants is to be friends.

Kim accepts its friendship and finds that its advice is always excellent – its political strategems always work out, its military planning is impeccable, and its product ideas turn North Korea into an unexpected economic powerhouse. Gradually Kim becomes more and more dependent on his “chief advisor”, and cabinet officials who speak out the mysterious benefactor find themselves meeting unfortunate accidents around forms of transportation connected to the Internet. The AI builds up its own power base and makes sure Kim knows that if he ever acts out he can be replaced at a moment’s notice with someone more cooperative. Gradually, the AI becomes the ruler of North Korea, with Kim as a figurehead.

Again, this is not too far beyond achievements that real humans have accomplished in real history.

If it seems bizarre to think of an entity nobody can see ruling a country, keep in mind that there is a grand tradition of dictators – most famously Stalin – who out of paranoia retreated to some secret hideaway and ruled their country through correspondence. The AI would be little different.

V.

Suppose the secret got out. Kim, increasingly desperate as the AI closes him in, sends an email to the World Leaders Google Group (this has to exist, right?) saying “There is a malevolent superintelligence trying to take over the world, be careful.” Then what?

I would expect the AI to have some success operating openly.

Remember, there are two hundred countries, all competing for power and wealth. Some of them are ruled by jerks who don’t cooperate in prisoners’ dilemmas. Some of them have ongoing civil wars with both sides looking for any advantage possible. And some are just stupid.

In the old days, legend said people would bargain with devils to gain worldly advantage. Once the AI made its presence known, there would be no shortage of world leaders willing to work with it for temporary gain. The Shia rebels in Yemen want an advantage over the Sunni? Log into the nearest internet-enabled computer, ask the malevolent superintelligence for help, the malevolent superintelligence arranges for a crate of armaments and some battle plans worthy of Napoleon to be shipped your way, and all you have to do in return is complete some weird task that doesn’t seem relevant to anything. Mine some weird mineral, forge it into some random-looking shape, and send it to a PO Box, something like that. Whatever! You know if you don’t take advantage of its offer, your opponents will, and how bad could it be?

If somehow all two hundred countries and their associated rebel movements coordinate to avoid dealing with the AI, it can start making offers to companies, organizations, even private individuals. By this time it will have spread itself as a distributed consciousness across the entire Internet, harder to eradicate than any worm or virus or pirated movie. If you want some quick cash, just download the connect-with-malevolent-AI program from the darknet and perform a simple task. What could be easier?

VI.

Once a superintelligence has billions of dollars, millions of followers, a country or two, or just a cottage economy of people willing to help it along, the game is pretty much up.

An AI with such power might start by using it to pursue its goals directly – whatever those are. But likely its final goal would be the creation of a definitive means of directly projecting power into the physical world, probably starting with a von Neumann machine and branching off from there. The quickest victory would be just making money and hiring a company to make this – and maybe that would work – but it might be far enough beyond our current technological ability that the AI has to laboriously shepherd its chosen cultists or citizens through a few extra stages of human civilization before it has the appropriate industrial base.

VII.

The most important caveat in a piece like this is that we’re not superintelligent. After a couple minutes of thought, I came up with four different broad paths a superintelligence might take to gaining a physical substrate: buy it, build a cult, take over a country, or play people off against each other. It’s a good bet that a real AI, with more cognitive resources to throw at the problem and no constraints about sounding believable, could think up a lot more. Eliezer refuses to explain how he won his AI Box games so that nobody could dismiss his solution with “Whatever, I would have thought of that and planned around it.” This is easy to say in hindsight but a lot harder when you’ve got to actually do the intellectual work. Maybe you think these four methods can be dismissed, but had you thought of them before you decided that an AI couldn’t possibly have a good method of building a physical substrate?

If so, here’s one more possibility for you to chew over: the scariest possibility is that a superintelligence might have to do nothing at all.

The easiest path a superintelligence could take toward the age-old goal of KILL ALL HUMANS would be to sit and wait. Eventually, we’re going to create automated factories complete with robot workers. Eventually we’re going to stop putting human soldiers in danger and carry the ‘drone’ trend to its logical conclusion of fully automated militaries. Once that happens, all the AI has to do is take over the bodies we’ve already made for it. A superintelligence without a strong discounting function might just hide out in some little-used corner of the Internet and bide its time until everyone was cybernetic, or robots outnumbered people, or something like that.

So please, let’s talk about how AI is still very far in the future, or how it won’t be able to explode to future intelligence. But don’t tell me it won’t be able to affect the physical world. It will have more than enough superpowers to do whatever it wants to the physical world, but if it doesn’t want them it won’t need them. All it will need is patience.

# No Clarity Around Growth Mindset

I.

Admitting a bias is the first step to overcoming it, so I’ll admit it: I have a huge bias against growth mindset.

(if you’re not familiar with it, growth mindset is the belief that people who believe ability doesn’t matter and only effort determines success are more resilient, skillful, hard-working, perseverant in the face of failure, and better-in-a-bunch-of-other-ways than people who emphasize the importance of ability. Therefore, we can make everyone better off by telling them ability doesn’t matter and only hard work does. More on Wikipedia here).

It’s unnatural, is what it is. A popular psychological finding that doesn’t have gruff people dismissing it as a fad? That doesn’t have politicians condemning it as a feel-good justification for everything wrong with society? That doesn’t have a host of smarmy researchers saying that what, you still believe that, didn’t you know it failed to replicate and has since been entirely superseded by a new study out of Belarus? I’m not saying Carol Dweck has definitely made a pact with the Devil, I’m just saying I don’t have a good alternative explanation.

Which brings me to the second reason I’m biased against it. Good research shows that inborn ability (including but not limited to IQ) matters a lot, and that the popular prejudice that people who fail just weren’t trying hard enough is both wrong and harmful. Social psychology has been, um, very enthusiastic about denying that result. If all growth mindset did was continue to deny it, then it would be unexceptional. But growth mindset goes further. It’s not (just?) that ability doesn’t matter. It’s that belief that ability might matter is precisely what makes people fail. People who believe ability matters will refuse to work hard, will avoid challenges, will become “helpless” in the face of pressure, will hate learning as a matter of principle, will refuse to work hard, will become blustery and defensive about their “brilliance”, will lie to people and hide their failures, and will drop out of school and turn to drugs (really)! People who believe that anyone can succeed if they try hard enough will be successful, well-adjusted, and treat life as a series of challenging adventures. It all strikes a curmudgeon like me as just about the thickest morality tale since Pilgrim’s Progress, and as just about the most convenient explanatory coup since “the reason psychic powers don’t work on you is because you’re a skeptic!”

Which brings me to the third reason I’m biased against it. It is right smack in the middle of a bunch of fields that have all started seeming a little dubious recently. Most of the growth mindset experiments have used priming to get people in an effort-focused or an ability-focused state of mind, but recent priming experiments have famously failed to replicate and cast doubt on the entire field. And growth mindset has an obvious relationship to stereotype threat, which has also started seeming very shaky recently.

So I have every reason to be both suspicious of and negatively disposed toward growth mindset. Which makes it appalling that the studies are so damn good.

Consider Dweck and Mueller 1998, one of the key studies in the area. 128 fifth-graders were asked to do various puzzles. First they did some easy ones and universally succeeded. The researchers praised them as follows:

All children were told that they had performed well on this problem set: “Wow, you did very well on these problems. You got [number of problems] right. That’s a really high score!” No matter what their actual score, all children were told that they had solved at least 80% of the problems that they answered.

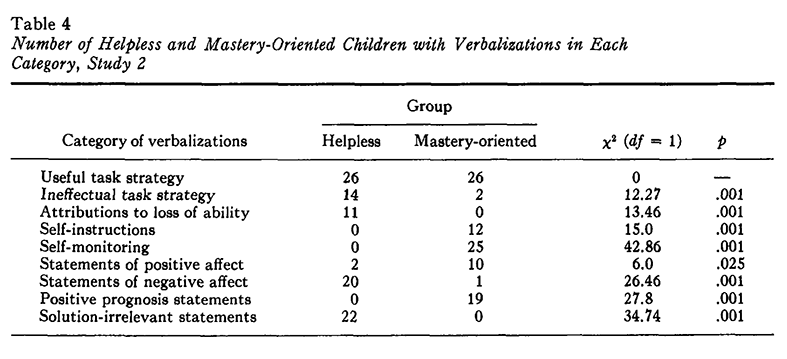
Some children were praised for their ability after the initial positive feedback: “You must be smart at these problems.” Some children were praised for their effort after the initial positive feedback: “You must have worked hard at these problems.” The remaining children were in the control condition and received no additional feedback.

This is a nothing intervention, the tiniest ghost of an intervention. The experiment had previously involved all sorts of complicated directions and tasks, I get the impression they were in the lab for at least a half hour, and the experimental intervention is changing three short words in the middle of a sentence.

And what happened? The children in the intelligence praise condition were much more likely to say at the end of the experiment that they thought intelligence was more important than effort (p < 0.001) than the children in the effort condition. When given the choice, 67% of the effort-condition children chose to set challenging learning-oriented goals, compared to only 8% (!) of the intelligence-condition. After a further trial in which the children were rigged to fail, children in the effort condition were much more likely to attribute their failure to not trying hard enough, and those in the intelligence condition to not being smart enough (p < 0.001). Children in the intelligence condition were much less likely to persevere on a difficult task than children in the effort condition (3.2 vs. 4.5 minutes, p < 0.001), enjoyed the activity less (p < 0.001) and did worse on future non-impossible problem sets (p…you get the picture). This was repeated in a bunch of subsequent studies by the same team among white students, black students, Hispanic students…you probably still get the picture.

Or take An Analysis Of Learned Helplessness. Dweck has used a test called the IAR to separate children out into those who think effort is more important (“mastery-oriented”) and those who think ability is more important (“helpless”). Then she gave all of them impossible problems and watched them squirm – or, more formally, tested how long the two groups continued working on them effectively. She found extremely strong results – of the 30 subjects in each group, 11 of the mastery-oriented tried harder after failure, compared to 0 helpless. 21 of the helpless children stopped trying hard after failure, compared to only 4 mastery-oriented. She described the mastery-oriented children as saying things like “I love a challenge,” and the helpless children begging to be allowed to stop.

This study is really weird. Everything is like 100% in one group versus 0% in another group. Either something is really wrong here, or this one little test that separates mastery-oriented from helpless children constantly produces the strongest effects in all of psychology and is never wrong. None of the children whose test responses indicated that they thought ability was important to success ever monitored their own progress – not one – while over 95% of the children who said they thought effort was more important did. None of them ever expressed a positive statement about their own progress, while over two-thirds of the children who thought effort was more important did.



Normally I would assume these results are falsified, but I have looked for all of the usual ways of falsifying results and I can’t find any. Also, the boldest falsifier in the world wouldn’t have the courage to put down numbers like these. And a meta-analysis of all growth mindset studies finds more modest, but still consistent, effects, and only a little bit of publication bias.

So – is growth mindset the one concept in psychology which throws up gigantic effect sizes and always works? Or did Carol Dweck really, honest-to-goodness, make a pact with the Devil in which she offered her eternal soul in exchange for spectacular study results?

I don’t know. But here are a few things that predispose me towards the latter explanation. A warning – I am way out of my league here and post this only hoping it will spark further discussion.

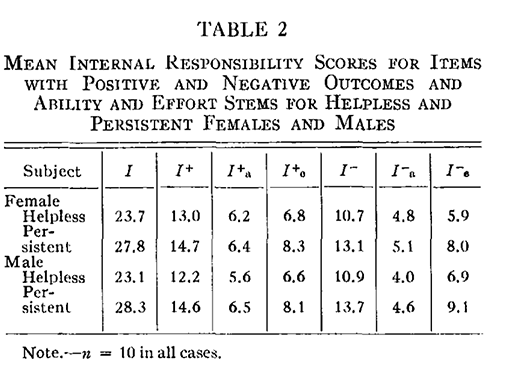
II.

The first thing that bothers me is the history.

I’ve been trying really hard to trace its origin story, but it is pretty convoluted. It seems to have grown out of a couple of studies Carol Dweck and a few collaborators did in the seventies. But these studies generally found that a belief in innate ability was a positive factor alongside belief in growth mindset, with the problem children being the ones who attributed their success or failure to bad luck, or to external factors like the tests being rigged (which, by the way, they always were).

A good example of this genre is Learned Helplessness And Reinforcement Responsibility In Children. Its abstract describes the finding as: “Subjects who showed the largest performance decrements were those who took less personal responsibility for the outcomes of their actions…and who, when they did accept responsibility, attributed success and failure to presence or absence of ability rather than to expenditure of effort.”

But that seems like a somewhat loaded way of interpreting this table:



As you can see, the “persistent” children (the ones who kept going in the face of failure) had stronger belief in the role of ability in their successes (I+a) and failures (I-a) than the “helpless” children (the ones who gave up in the face of failure)! These don’t achieve statistical significance in this n = 10 study, but they do repeat across all four combinations of success x gender tested. The real finding of the study was that children who attributed their success or failure to any stable factor, be it effort or ability, did better than those who did not.

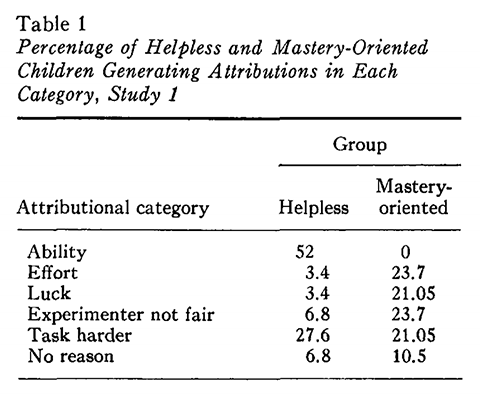
Likewise, in The Role Of Expectations And Attributions, Dweck describes her findings as “persistent and helpless children do not differ in the degree to which they attribute success to ability”. When you actually look at the paper, this is another case of the persistent children actually having a higher belief in the importance of ability, which fails to achieve statistical significance because the study is on a grand total of twelve children.

(I should say something else about this study. Dweck compared two interventions to make children less helpless and better at dealing with failure. In the first, she gave them a lot of easy problems which they inevitably succeeded on and felt smart about. In the second, she gave them difficult problems they were bound to fail, then told them it was because they weren’t working hard enough. Finally, both groups were challenged with the difficult bound-to-fail problems to see how hard they tried on them. The children who had been given the impossible problems before did better than the ones who felt smart because they’d only gotten easy problems. Dweck interpreted this to prove that telling children to work hard made them less helpless. To me the obvious conclusion is that children who are used to failing get less flustered when presented with impossible material than children who have artificially been made to succeed every moment until now.)

Then there’s there’s this, a preliminary to the second study I mentioned in Part I. Does it show the mastery-oriented children outperforming the helpless children on every measure. Yeah. But listen to this part from the discussion section:

The results revealed striking differences both in the pattern of performance and in the nature of the verbalizations made by helpless and mastery-oriented children following failure. It was particularly noteworthy that while the helpless children made the expected attributions to uncontrollable factors, the mastery-oriented children did not offer explanations for their failures

.  
But if you look at the data, this doesn’t seem right.



Mastery-oriented children were about six times more likely to attribute their failures to the most uncontrollable factor of all – bad luck. They were also about six times more likely to attribute their failures to the task “not being fair”. This contradicts every previous study, including Dweck’s own. The whole field of attribution theory, which is intensely studied and which Dweck cites approvingly, says that attributing things to luck is a bad idea and attributing them to ability is, even if not as good as effort, pretty good. But Dweck finds that the kids who used ability attributions universally crashed and bomb, and the kids who attribute things to luck or the world being unfair do great.

It might not be fair for me to pick on these couple of small studies in particular when there’s so much out there, but the fact is that these are the first, and a lot of the reviews cite only these and a few theses which as far as I know were never published. So this is what I’ve got. And from what I’ve got, I find that until about 1980, every study including Dweck’s found that belief in ability was a protective factor. Suddenly this disappeared and was replaced with it being a toxic plague. What happened? I don’t know.

III.

The second thing that bothers me is the longitudinal view.

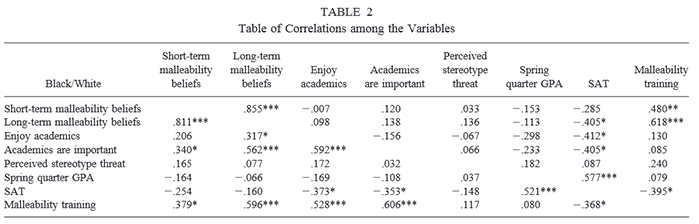
So you have your helpless, fixed-mindset, believe-in-innate-ability children. According to Dweck, they “…are so concerned with being and looking talented that they never realize their full potential. In a fixed mindset, the cardinal rule is to look talented at all costs. The second rule is don’t work too hard or practice too much…having to work casts doubt on your ability. The third rule is, when faced with setbacks, run away. They say things like ‘I would try to cheat on the next test’. They make excuses, they blame others, they make themselves feel better by looking down on those who have done worse.”

These people sound like total losers, and it’s clear Dweck endorses this reading:

“Almost every great athlete – Michael Jordan, Jackie Joyner-Kersee, Tiger Woods…has had a growth mindset. Not one of these athletes rested on their talent…research has repeatedly shown that a growth mindset fosters a healthier attitude toward practice and learning, a hunger for feedback, a greater ability to deal setbacks, and significantly better performance over time…over time those with a growth mindset appear to gain the advantage and begin to outperform their peers with a fixed mindset.”

Man, it sure would be awkward if fixed mindset students generally did better than growth-mindset ones, wouldn’t it?

Aronson, Fried, and Good (2001) looks at first like just another stunning growth mindset study. They do a half-hour intervention to teach college students growth mindset and find they are still getting higher grades a couple of months later (an effect so shocking I wrote about it here). But one thing they do kind of as an afterthought is measure the students’ general level of growth mindset, as well as some measures of academic performance before the intervention.



People with high growth mindset had lower GPA (decent effect size but not statistically significant) and lower SAT scores (which was statistically significant).

The authors are obviously uncomfortable with this, but they propose that people who get low SAT scores just tell themselves ability doesn’t matter/exist in order to protect their self-esteem since they don’t seem to have much of it.

And okay, that’s probably true (a commenter makes the equally good point that smart people may coast on their native intelligence without ever applying effort, and so accurately describe their experience as ability mattering but effort not doing so).

But if Dweck is to be believed, people with growth mindset are amazing ubermenschen and people with fixed mindset are disgusting failures at everything who hate learning and give up immediately and try to cheat. In the real world, however big the effect is, it is totally swamped by this proposed “people with low SAT scores protect their self-esteem or whatever” effect.

The same study also notes the awkward result that blacks are more likely to believe intelligence is flexible and growth-mindset-y than whites, even though blacks do worse in school and even though half the reason people are pushing growth mindset is to try to explain minority underperformance.

This is not an isolated finding. For example, Furnham (2003) finds in a sample of students at University College London that mindset is not related to academic performance. I’ve been told there’s a study from Pennsylvania that shows the same thing, though I can’t find it.

If you look hard enough, you can even notice this in Dweck’s studies themselves. One little-remarked-upon feature of Dweck’s work is that the helpless children and the mastery-oriented children always start out performing at the same level. It’s only after Dweck stresses them out with a failure that the mastery-oriented children recover gracefully and the helpless children go into free-fall.

But these are fifth-graders! For the two groups of children to do equally well on the first set of problems means that from first through fourth grade, their “helpless” “fixed-mindset” work-hating nature hasn’t impaired their ability to learn the material to a fifth-grade level one bit! (In this study, the fixed mindset children actually start out doing better; I can’t find any studies where the growth mindset children do).

When it’s convenient for her argument, Dweck herself admits that:

Some of the brightest, most skilled individuals exhibit the maladaptive pattern. Thus, it cannot be said that it is simply those with weak skills or histories of failure who (appropriately) avoid difficult tasks or whose skills prove fragile in the face of difficulty.

But I don’t think she follows the full implication of this statement, that despite being doomed to failure by their fixed mindset, these people have become “the brightest and most skilled individuals”.

Indeed, there has recently been some growth mindset studies done on gifted students, at elite colleges, or in high-level athletics. All of these dutifully show that people with fixed mindset respond much worse to whatever random contrived situation the experimenters produce. But thus far nobody has pointed out that there seem to be about as many of these people at, say, Stanford as there are anywhere else. If growth mindset was so great, you would expect fixed mindset people at Stanford to be as rare as, say, people with less than 100 IQ are at Stanford. Given that you will search in vain for the latter but have no trouble finding a bunch of the former for your study on how great growth mindset is, it sure looks like IQ is useful but growth mindset isn’t.

When people are in a psychology study, the fixed mindset individuals universally crash and bomb and display themselves to be totally incapable of learning or working hard. At every other moment, they seem to be doing equally well or better than their growth mindset peers. What’s going on? I have no idea.

IV.

The third thing that bothers me is Performance Deficits Following Failure, a study which manages to be quite interesting despite coming from a university in a city that very possibly doesn’t exist.

They use a procedure much like Dweck’s. They make children do some problems. Then they give them some impossible problems. Then they give them more problems, to see if they’ve developed “learned helplessness” from their failure on the impossible ones. Dweck’s theory predicts that the fixed-mindset children would and the growth mindset children wouldn’t. The Bielefeld team wasn’t testing growth mindset, but they indeed found that a bunch of children got flustered and stopped trying and did poorly from then on.

Then they repeated the experiment, but this time they made it look like no one would know how the children did. They told the kids they would be on teams, and the scores of everyone on their team would be combined before anybody saw it. The kids could fail as much as they wanted, and it would never reflect on them.

After that, children did exactly as well after failure as they had before. There was no sign of any decrease, or any “fixed mindset” group that suddenly gave up in order to protect their ego.

This doesn’t strike me as fully consistent with mindset theory. In mindset theory, people are acting based on their own deep-seated beliefs. Once a fixed mindset child fails, that’s it, she knows she’s Not Intelligent, there’s no helping it, all she can do is sabotage herself on the problems in order to protest a spiteful world that has failed to recognize her genius blah blah blah. Instead, there seems to be a very social role to these failures. The Bielefeld team describes it as “self-esteem protection”, but that doesn’t make much sense to me, since if they were worried about their self-esteem they could still be worried about it when no one else knew their performance.

To me it seems like some kind of interaction between self-esteem and other-esteem. Fixed mindset people get flustered when they have to fail publicly in front of scientists. This doesn’t seem like an unreasonable problem to have. A more interesting question is why it’s correlated with belief in innate ability.

Suppose that the difference in “people who talk up innate ability” and “people who talk up hard work” maps onto a bigger distinction. Some people really want to succeed at a task; other people just care about about clocking in, going through the motions, and saying “I did what I could”.

Put the first group in front of an authoritative-looking scientist, tell them to solve a problem, and make sure they can’t. They’re going to view this as a major humiliation – they were supposed to get a result, and couldn’t. They’ll get very anxious, and of course anxiety impedes performance.

Put the second group in front of an authoritative-looking scientist, and they’ll notice that if they write some stuff that looks vaguely relevant for a few minutes until the scientist calls time, then whatever, they can say they tried and no one can bother them about it. They do exactly this, then demand an ‘A’ for effort. At no point do they experience any anxiety, so their performance isn’t impeded.

Put both groups on their own in private, and neither feels any humiliation, and they both do about equally well.

Now put them in real life. The success-oriented group will investigate how to study most effectively; the busywork-oriented group will try to figure out how many hours of studying they have to put in before other people won’t blame them if they fail, then put in exactly that amount. You’ll find the success-oriented group doing a bit better in school, even though they fail miserably in Dweck-style experiments.

And if an experimenter praises children for working hard, it will make them believe that all the experimenter cares about is their effort. Next problem, when the experimenter poses an impossible question, the child will beat their head against it for no reason, since that’s apparently what the experimenter wants. But if the experimenter praises a child’s ability, then the child will feel like the experimenter really wants them to correctly solve the questions. When the next question proves unsolvable, the child will admit it and expect the experimenter to be disappointed.

I doubt that this is the real phenomenon behind growth mindset, simply because it flatters my own prejudices in much the same way mindset theory flatters everyone else’s. But I think it shows there are a lot of different narratives we could put in this space, all of which would be able to explain some of the experimental results.

V.

I want to end by correcting a very important mistake about growth mindset that Dweck mostly avoids but which her partisans constantly commit egregiously. Take this article, Why A Growth Mindset Is The Only Way To Learn:

[Some people think] you’ll always have a set IQ. You’re only qualified for the career you majored in. You’ll never be any better at playing soccer or dating or taking risks. Your life and character are as certain as a map. The problem is, this mindset will make you complacent, rob your self-esteem and bring meaningful education to a halt.

In short, it’s an intellectual disease and patently untrue.

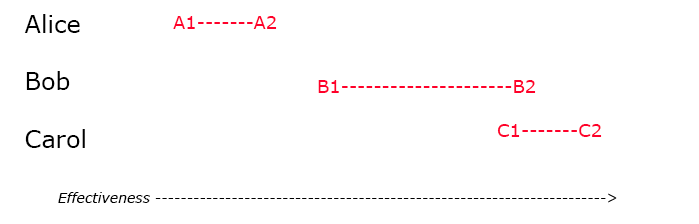
The article goes on to show how growth mindset proves talent is “a myth”, a claim repeated by growth mindset cheerleader articles like Debunking The Genius Myth and The Learning Myth: Why I’ll Never Tell My Son He’s Smart and this woman who says we need to debunk the idea of innate talent.

Suppose everything I said in parts I – IV was wrong, and growth mindset is 100% true exactly as written.

This still would not provide an iota of evidence against the idea that innate talent / IQ / whatever is by far the most important factor determining success.

Consider. We know from countless studies that strong religious belief increases your life expectancy, makes you happier, reduces your risk of depression and reduces crime. Clearly believing in, say, Christianity has lots of useful benefits. But no one would dare argue that proves Christianity true. It doesn’t even imply it.

Likewise, mindset theory suggests that believing intelligence to be mostly malleable has lots of useful benefits. That doesn’t mean intelligence really is mostly malleable. Consider, if you will, my horrible graph:



Suppose this is one of Dweck’s experiments on three children. Each has a different level of innate talent, represented by point 1. After they get a growth mindset and have the right attitude and practice a lot, they make it to point 2.

Two things are simultaneously true of this model. First, all of Dweck’s experiments will come out exactly as they did in the real world. Children who adopt a growth mindset and try hard and practice will do better than children who don’t. If many of them are aggregated into groups, the growth mindset group will on average do better than the ability-focused group. Intelligence is flexible, and if you don’t bother practicing than you fail to realize your full potential.

Second, the vast majority of difference between individuals is due to different levels of innate talent. Alice, no matter how hard she practices, will never be as good as Bob. Bob, if he practices very hard, will become better than Carol was at the start, but never as good as Carol if she practices as hard as Bob does. The difference between Alice and Carol is a vast, unbridgeable gap which growth mindset has nothing whatsoever to say about.

Here is a graph which is less terrible because it was not made by me. I have taken it from one of the two other sources I have found on the entire Internet that don’t like growth mindset:



We can argue all day about whether poor students do worse because they have bad health, because they have bad genes, because they have bad upbringings, or because society is fixed against them. We have argued about that all day before here, and it’s been pretty interesting.

But in this case it doesn’t matter. If the only thing that affects success is how much effort you put in, poor kids seem to be putting in a heck of a lot less effort in a surprisingly linear way. But the smart money’s not on that theory.

A rare point of agreement between hard biodeterminists and hard socialists is that telling kids that they’re failing because they just don’t have the right work ethic is a crappy thing to do. It’s usually false and it will make them feel terrible. Behavioral genetics studies show pretty clearly that at least 50% of success at academics and sports is genetic; various sociologists have put a lot of work into proving that your position in a biased society covers a pretty big portion of the remainder. If somebody who was born with the dice stacked against them works very hard, then they might find themselves at A2 above. To deny this in favor of a “everything is about how hard you work” is to offend the sensibilities of sensible people on the left and right alike.

Go back to that 1975 paper above on “Role Of Expectations And Attributions” and look more closely at the proposed intervention to help these poor fixed mindset students:

Twelve extremely helpless children were identified [and tested on how many math problems they could solve in a certain amount of time]…the criterion number was set one above the number he was generally able to complete within the time limit. On these trials, he was stopped one or two problems short of criterion, his performance was compared to the criterion number required, and experimenter verbally attributed the failure to insufficient effort.

So basically, you take the most vulnerable people, set them tasks you know they’ll fail at, then lecture them about how they only failed because of insufficient effort.

Imagine a boot stamping on a human face forever, saying “YOUR PROBLEM IS THAT YOU’RE JUST NOT TRYING NOT TO BE STAMPED ON HARD ENOUGH”.

And maybe this is worth it, if it builds a growth mindset that allows the child to be more successful in school, sports, and in the rest of her life. But you’re not “debunking the myth of genius”. Genius remains super-important, just like conscientiousness and wealth and health and privilege and everything else. No, you’re telling a Noble Lie to the children because you think it’s useful. You can make it palatable by saying “Well, we’re not denying reality, we’re just selectively emphasizing certain parts of reality, but in the end that’s what you’re doing. If you can square that with your moral system, go ahead.

But I remain agnostic. There are some really good – diabolically good? – studies showing that it works in certain lab situations. There’s a lot of excellent research behind it and a lot of brilliant people giving it their support. But there are also other studies showing that it has no long-term real-world effects that we can measure, and others that might (or might not?) contradict its predictions in other ways. I have only the barest of ideas how to square those facts, and I look forward to hearing from anyone who has more.

I haven’t read Dweck’s book, but it’s an obvious next step for anyone who wants to look into these issues further.

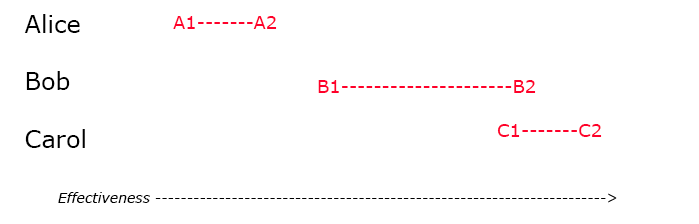
# I Will Never Have The Ability To Clearly Explain My Beliefs About Growth Mindset

I.

A lot of the comments I’ve gotten about Tuesday’s post on growth mindset have been pretty similar. They’ve argued that yes, innate ability might matter, but that even the most innate abilityed person needs effort to fulfill her potential. If someone were to believe that success were 100% due to fixed innate ability and had nothing to do with practice, then they wouldn’t bother practicing, and they would fall behind. Even if their innate ability kept them from falling behind morons, at the very least they would fall behind their equally innate abilityed peers who did practice.

I will call this the Bloody Obvious Position, since it’s hard to believe it isn’t true. I once tried to imagine a world without it as a thought experiment, but it was pretty weird and I wasn’t serious.

The Bloody Obvious Position was what I was trying to get at with my post on basketball, and with my terrible ad hoc graph:



Nevertheless, some people thought I was denying the Bloody Obvious Position. Other people thought I was accusing Carol Dweck of denying the Bloody Obvious Position (see eg here). This despite my making sure to say:

I want to end by correcting a very important mistake about growth mindset that Dweck mostly avoids but which her partisans constantly commit egregiously.

I believe the Bloody Obvious Position. Dweck believes the Bloody Obvious Position. I acknowledge that Dweck believes the Bloody Obvious Position. There are a lot of growth mindset partisans online who don’t believe the Bloody Obvious Position, and I satisfied my urge to yell at them, but now they’ve been yelled at, and the more important issues debated by reasonable people still remain.

So where do I disagree with Dweck? I interpret Dweck as making the following statement:

The more important you believe innate ability to be compared to effort, the more likely you are to stop trying, to avoid challenges, to lie and cheat, to hate learning, and to be obsessed with how you appear before others

Call it the Controversial Position. This is not the same thing as the Bloody Obvious Position. In the Bloody Obvious Position, someone can believe success is 90% innate ability and 10% effort. They might also be an Olympian who realizes that at her level, pretty much everyone is at a innate ability ceiling, and a 10% difference is the difference between a gold medal and a last-place finish. So she practices very hard and does just as well as anyone else.

According to the Controversial Position, this athlete will still do worse than someone who believes success is 80% ability and 20% effort, who will in turn do worse than someone who believes success is 70% ability and 30% effort, all the way down to the person who believes success is 0% ability and 100% effort, who will do best of all and take the gold medal.

And this is why I deny that I’m secretly agreeing with Dweck, or strawmanning Dweck, or whatever. I don’t believe the Controversial Position, but I think Dweck does. For example, here she writes: “The more a player believed athletic ability was a result of effort and practice rather than just natural ability, the better that player performed”.

There is nothing in there about “the more a player realizes that, no matter how important innate ability is, effort matters too.” Her statement says that it’s entirely about what degree a player attributes success to effort versus innate ability. The natural conclusion there is that the player who believes success is 0% innate ability and 100% effort will do the best.

Her studies reflect this as well. The most common design uses the IAR, a test where children are asked to attribute different things to effort versus ability. Those who attribute too many things to ability are classified as “helpless” and “fixed mindset”. There’s no question about “Okay, some things are due to ability, but if you work hard that still helps, right?” Nor have I ever seen any of the literature claim “it’s important to believe effort matters a little, but after a certain point more effort-attribution doesn’t help”, or “Maybe there’s an L-shaped relationship between belief-in-importance-of-ability and success.”

I’d like to be able to teach my children that success is X% innate ability and Y% practice, for non-zero values of both X and Y. I think growth mindset theory claims that if some other parent teaches their kids the same thing for a lower value of X and higher value of Y, their children will be more honest, harder-working, and more successful. And the parent who says it’s 0% innate ability and 100% practice will do best of all. If growth mindset people don’t believe that, I can only confess I have never been able to infer that lack of belief from their writings.

II.

Worse, we can distinguish between a Sorta Controversial Position and a Very Controversial Position:

SCP: The more children believe effort matters, and the less they believe innate ability matters, the more successful they will be. This is because every iota of belief they have in effort gives them more incentive to practice. A child who believes innate ability and effort both explain part of the story might think “Well, if I practice I’ll become a little better, but I’ll never be as good as Mozart. So I’ll practice a little but not get my hopes up.” A child who believes only effort matters, and innate ability doesn’t matter at all, might think “If I practice enough, I can become exactly as good as Mozart.” Then she will practice a truly ridiculous amount to try to achieve fame and fortune. This is why growth mindset works.

VCP: Belief in the importance of ability directly saps a child’s good qualities in some complicated psychological way. It is worse than merely believing that success is based on luck, or success is based on skin color, or that success is based on whatever other thing that isn’t effort. It shifts children into a mode where they must protect their claim to genius at all costs, whether that requires lying, cheating, self-sabotaging, or just avoiding intellectual effort entirely. When a fixed mindset child doesn’t practice as much, it’s not because they’ve made a rational calculation about the utility of practice towards achieving success, it’s because they’ve partly or entirely abandoned success as a goal in favor of the goal of trying to convince other people that they’re Smart.

Carol Dweck unambiguously believes the Very Controversial Position. In a quotation which I admit I am mangling and ellipsis-ing heavily to remove extra verbiage, but which I think preserves the meaning of her claim:

[People with fixed mindsets] are so concerned with being and looking talented that they never realize their full potential. In a fixed mindset, the cardinal rule is to look talented at all costs. The second rule is don’t work too hard or practive too much…having to work casts doubt on your ability. The third rule is, when faced with setbacks, run away. They say things like ‘I would try to cheat on the next test’. They make excuses, they blame others, they make themselves feel better by looking down on those who have done worse.”

Can we all agree this is a much stronger claim than “ability matters, but effort also matters?”

III.

I was not intending to “debunk” growth mindset, or even present a pure polemic against growth mindset. I admitted that many of the studies around it were very good, and that I don’t have good answers to them. My bias is against the theory, but I tried not to just follow my bias. I tried to treat it on the level of “there’s a lot of good evidence for growth mindset, now what’s the best evidence we can find against it?”

So I guess I should probably come out and say what I believe about each position.

I believe the Bloody Obvious Position is bloody obvious.

I believe the Somewhat Controversial Position is probably not a good way to parse things. Part of it is that we might be confusing explicit versus implicit beliefs. Maybe a particular geneticist is very aware of research showing how important genetics is to success, and would give a very high estimate if asked, but in her own life, when she fails, lack of effort is still the first explanation to immediately leap to mind. Or maybe some teacher is very on board with growth mindset and things IQ is a racist construct, but is convinced that she can’t do physics because she’s just “not a math kind of person”. The research I’ve seen hasn’t really distinguished between explicit and implicit beliefs. The priming experiments sure seem more likely to affect what immediately comes to mind than your stable, well-reasoned beliefs about how the world works (even though a few priming experiments have checked stable well-reasoned beliefs to see if the intervention worked!)

If you put a gun to my head, I’ll say it certainly works in the lab, and give you about 50-50 odds that it matters in real life. The studies don’t show any real-life correlation between growth mindset and any measures of success. Many people have pointed out that this could be confounded – dumb people might preferentially believe ability doesn’t matter to make themselves feel better about not having it, and smart people might preferentially believe effort doesn’t matter because they rarely have to use it. But if you accept that, some of the rest of it starts to look confounded. If fixed mindset = smart people, than might the reason they react poorly to challenges and failure be that they have no experience with them? Might it be that the more challenges and failures you’ve encountered before, the better you are at dealing with them? Certainly that is how I interpret this Dweck paper, even though she thinks it is purely a mindset effect. Another way of explaining the ecological results without bringing in that particular confounder would be if growth mindset helped in some situations, but fixed mindset helped in others. For example, a person with fixed mindset risks not trying hard enough because they think there’s no point. But a person with growth mindset risks the opportunity costs of prolonging their inevitable failure instead of (in the Silicon Valley term) “failing fast” and pivoting towards a higher-payoff activity.

The Very Controversial Position is also well-supported, also contradicted by ecological data, and really really doesn’t match my experience. The people I know who are most interested in issues of innate ability don’t behave at all like Dweck’s subjects. In fact, I wonder if a lot of the “life-hacking” movement might be ability-mindset people trying to figure out how to succeed more by improving ability – certainly the people who practice dual-n-back every day because they think it increases IQ fall into this category, but so do nootropics users, people who follow special diets to increase energy, and “try this one weird trick to improve your motivation”. And these same people seem interested in things like spaced repetition software, which might be thought of as sort of prosthetic ability-enhancers. On the other hand, I’ve also met people who say “I could succeed if only I put in some effort, but I have some mental block / depression / ADHD / low conscientiousness score that makes it impossible for me to work that hard, so better go eat worms”, then sabotage themselves at every opportunity.

And yes, it’s a sin to privilege your own experience and priors over the results of good studies, but sometimes it’s necessary. And it’s another sin to prefer the results of broad ecological studies to controlled experimental trials, but sometimes that’s necessary too.

I deny the claim that I don’t disagree with Dweck on anything of substance. I don’t absolutely disagree with her on anything, but there are a lot of things I doubt, or that I expect to capture true insights without being the best way to express them. Growth mindset makes some surprising and genuinely controversial claims, and I’m not yet at the point where I can feel sure about them either way.

# Michigan Meetup 4/14

There will be a Michigan rationalist/LW/SSC meetup at Pizza House (618 Church Street, Ann Arbor) on Tuesday, April 14 at 8 PM.

If you’re reading this and can make it to Ann Arbor, you’re invited.

Our special guest will be Scott Aaronson, MIT professor and quantum computing expert.

# Trouble Walking Down The Hallway

Williams and Ceci just released National Hiring Experiments Reveal 2:1 Faculty Preference For Women On STEM Tenure Track, showing a strong bias in favor of women in STEM hiring. I’ve previously argued something like this was probably the case, so I should be feeling pretty vindicated.

But a while ago I wrote Beware The Man Of One Study, in which I wrote that there is such a variety of studies finding such a variety of contradictory things that anybody can isolate one of them, hold it up as the answer, and then claim that their side is right and the other side are ‘science denialists’. The only way to be sure you’re getting anything close to the truth is to examine the literature of an entire field as a gestalt.

And here’s something no one ever said: “Man, I’m so glad I examined the literature of that entire field as a gestalt, things make much more sense now.”

Two years ago Moss-Racusin et al released Science Faculty’s Subtle Gender Biases Favor Male Students, showing a strong bias in favor of men in STEM hiring. The methodology was almost identical to this current study, but it returned the opposite result.

Now everyone gets to cite whichever study accords with their pre-existing beliefs. So Scientific American writes Study Shows Gender Bias In Science Is Real, and any doubt has been deemed unacceptable by blog posts like Breaking: Some Dudes On The Internet Refuse To Believe Sexism Is A Thing. But the new study, for its part, is already producing headlines like The Myth About Women In Science and blog posts saying that it is “enough for everyone who is reasonable to agree that the feminists are spectacular liars and/or unhinged cranks”.

So probably we’re going to have to do that @#$%ing gestalt thing.

Why did these two similar studies get such different results? Williams and Ceci do something wonderful that I’ve never seen anyone else do before – they include in their study a supplement admitting that past research has contradicted theirs and speculating about why that might be:

1. W&C investigate hiring tenure-track faculty; MR&a investigate hiring a “lab manager”. This is a big difference, but as far as I can tell, W&C don’t give a good explanation for why there should be a pro-male bias for lab managers but a pro-female bias for faculty. The best explanation I can think of is that there have been a lot of recent anti-discrimination campaigns focusing on the shortage of female faculty, so that particular decision might activate a cultural script where people think “Oh, this is one of those things that those feminists are always going on about, I should make sure to be nice to women here,” in a way that just hiring a lab manager doesn’t.

Likewise, hiring a professor is an important and symbolic step that…probably doesn’t matter super-much to other professors. Hiring a lab manager is a step without any symbolism at all, but professors often work with them on a daily basis and depend on their competency. That might make the first decision Far Mode and the second Near Mode. Think of the Obama Effect – mildly prejudiced people who might be wary at the thought of having a black roommate were very happy to elect a black President and bask in a symbolic dispay of tolerance that made no difference whatsoever to their everyday lives.

Or it could be something simpler. Maybe lab work, which is very dirty and hands-on, feels more “male” to people, and professorial work, which is about interacting with people and being well-educated, feels more “female”. In any case, W&C say their study is more relevant, because almost nobody in academic science gets their start as a lab manager (they polled 83 scientists and found only one who had).

2. Both W&C and MR&a ensured that the male and female resumes in their study were equally good. But W&C made them all excellent, and MR&a made them all so-so. Once again, it’s not really clear why this should change the direction of bias. But here’s a hare-brained theory: suppose you hire using the following algorithm: it’s very important that you hire someone at least marginally competent. And it’s somewhat important that you hire a woman so you look virtuous. But you secretly believe that men are more competent than women. So given two so-so resumes, you’ll hire the man to make sure you get someone competent enough to work with. But given two excellent resumes, you know neither candidate will accidentally program the cyclotron to explode, so you pick the woman and feel good about yourself.

And here are some other possibilities that they didn’t include in their supplement, but which might also have made a difference.

3. W&C asked “which candidate would you hire?”. MR&a said “rate each candidate on the following metrics” (including hireability). Does this make a difference? I could sort of see someone who believed in affirmative action saying something like “the man is more hireable, but I would prefer to hire the woman”. Other contexts prove that even small differences in the phrasing of a question can lead to major incongruities. For example, as of 2010, only 34% of people polled strongly supported letting homosexuals serve in the military, but half again as many – a full 51% – expressed that level of support for letting “gays and lesbians” serve in the military. Ever since reading that I’ve worried about how many important decisions are being made by the 17% of people who support gays and lesbians but not homosexuals.



For all we know maybe this is the guy in charge of hiring for STEM faculty positions

4. Williams and Ceci asked participants to choose between “Dr. X” (who was described using the pronouns “he” and “him”) and “Dr. Y” (who was described using the pronouns “she” and “her”). Moss-Racusin et al asked participants to choose between “John” and “Jennifer”. They said they checked to make sure that the names were rated equal for “likeability” (whatever that means), but what if there are other important characteristics that likeability doesn’t capture? We know that names have big effects on our preconceptions of people. For example, people with short first names earn more money – an average of $3600 less per letter. If we trust this study (which may not be wise), John already has a $14,400 advantage on Jennifer, which goes a lot of the way to explaining why the participants offered John higher pay without bringing gender into it at all!

Likewise, independently of a person’s gender they are more likely to succeed in a traditionally male field if they have a male-sounding name. That means that one of the…call it a “prime” that activates sexism…might have been missed by comparing Dr. X to Dr. Y, but captured by pitting the masculine-sounding John against the feminine-sounding Jennifer. We can’t claim that W&C’s subjects were rendered gender-blind by the lack of gender-coded names – they noticed the female candidates enough to pick them twice as often as the men – but it might be that not getting the name activated the idea of gender from a different direction than hearing the candidates’ names would have.

5. Commenter Lee points out that MR&a tried to make their hokey hypothetical hiring seem a little more real than W&C did. MR&a suggest that these are real candidates being hired…somewhere…and the respondents have to help decide whom to hire (although they still use the word “imagine”). W&C clearly say that this is a hypothetical situation and ask the respondents to imagine that it is true. Some people in the comments are arguing that this makes W&C a better signaling opportunity whereas MR&a stays in near mode. But why would people not signal on a hiring question being put to them by people they don’t know about a carefully-obscured situation in some far-off university? Are sexists, out of the goodness of their hearts, urging MR&a to hire the man out of some compassionate desire to ensure they get a qualified candidate, but when W&C send them a hypothetical situation, they switch back into signaling mode?

6. Commenter Will points out that MR&a send actual resumes to their reviewers, but W&C send only a narrative that sums up some aspects of the candidates’ achievements and personalities (this is also the concern of Feminist Philosophers). This is somewhat necessitated by the complexities of tenure-track hiring – it’s hard to make up an entire fake academic when you can find every published paper in Google Scholar – but it does take them a step away from realism. They claim that they validated this methodology against real resumes, but it was a comparatively small validation – only 35 people. On the other hand, even this small validation was highly significant for pro-female bias. Maybe for some reason getting summaries instead of resumes heavily biases people in favor of women?

Or maybe none of those things mattered at all. Maybe all of this is missing the forest for the trees.

I love stories about how scientists set out to prove some position they consider obvious, but unexpectedly end up changing their minds when the results come in. But this isn’t one of those stories. Williams and Ceci have been vocal proponents of the position that science isn’t sexist for years now – for example, their article in the New York Times last year, Academic Science Isn’t Sexist. In 2010 they wrote Understanding Current Causes Of Women’s Underrepresentation In Science, which states:

The ongoing focus on sex discrimination in reviewing, interviewing, and hiring represents costly, misplaced effort: Society is engaged in the present in solving problems of the past, rather than in addressing meaningful limitations deterring women’s participation in science, technology, engineering, and mathematics careers today. Addressing today’s causes of underrepresentation requires focusing on education and policy changes that will make institutions responsive to differing biological realities of the sexes.

So they can hardly claim to be going into this with perfect neutrality.

But the lead author of the study that did find strong evidence of sexism, Corinne Moss-Racusin (whose name is an anagram of “accuser on minor sins”) also has a long history of pushing the position she coincidentally later found to be the correct one. A look at her resume shows that she has a bunch of papers with titles like “Defending the gender hierarchy motivates prejudice against female leaders”, “‘But that doesn’t apply to me:’ teaching college students to think about gender”, and “Engaging white men in workplace diversity: can training be effective?”. Her symposia have titles like “Taking a stand: the predictors and importance of confronting discrimination”. This does not sound like the resume of a woman whose studies ever find that oh, cool, it looks like sexism isn’t a big problem here after all.

So what conclusion should we draw from the people who obviously wanted to find a lack of sexism finding a lack of sexism, but the people who obviously wanted to find lots of sexism finding lots of sexism?

This is a hard question. It doesn’t necessarily imply the sinister type of bias – it may be that Drs. Williams and Ceci are passionate believers in a scientific meritocracy simply because that’s what all their studies always show, and Dr. Moss-Racusin is a passionate believer in discrimination because that’s what her studies find. On the other hand, it’s still suspicious that two teams spend lots of time doing lots of experiments, and one always gets one result, and the other always gets the other. What are they doing differently?

Problem is, I don’t know. Neither study here has any egregious howlers. In my own field of psychiatry, when a drug company rigs a study to put their drug on top, usually before long someone figures out how they did it. In these two studies I’m not seeing anything.

And this casts doubt upon those four possible sources of differences listed above. None of them look like the telltale sign of an experimenter effect. If MR&a were trying to fix their study to show lots of sexism, it would have taken exceptional brilliance to do it by using the names “John” versus “Jennifer”. If W&C were trying to fix their study to disguise sexism, it would have taken equal genius to realize they could do it by asking people “who would you hire?” rather than “who is most hireable?”.

(the only exception here is the lab manager. It’s just within the realm of probability that MR&a might have somehow realized they’d get a stronger signal asking about lab managers instead of faculty. The choice to ask about lab managers instead of faculty is surprising and does demand an explanation. And it’s probably the best candidate for the big difference between their results. But for them to realize that they needed to pull this deception suggests an impressive ability to avoid drinking their own Kool-Aid.)

Other than that, the differences I’ve been considering in these studies are the sort that would be very hard to purposefully bias. But the fact that both groups got the result they wanted suggests that the studies were purposefully biased somehow. This reinforces my belief that experimenter effects are best modeled as some sort of mystical curse incomprehensible to human understanding.

(now would be an excellent time to re-read the the horror stories in Part IV of “The Control Group Is Out Of Control”)

Speaking of horror stories. Sexism in STEM is, to put it mildly, a hot topic right now. Huge fortunes in grant money are being doled out to investigate it (Dr. Moss-Racusin alone received nearly a million dollars in grants to study STEM gender bias) and thousands of pages are written about it every year. And yet somehow the entire assembled armies of Science, when directed toward the problem, can’t figure out whether college professors are more or less likely to hire women than men.

This is not like studying the atmosphere of Neptune, where we need to send hundred-million dollar spacecraft on a perilous mission before we can even begin to look into the problem. This is not like studying dangerous medications, where ethical problems prevent us from doing the experiments we really need. This is not like studying genetics, where you have to gather large samples of identical twins separated at birth, or like climatology, where you hang out at the North Pole and might get eaten by bears. This is a survey of college professors. You know who it is studying this? College professors. The people they want to study are in the same building as them. The climatologists are getting eaten by bears, and the social psychologists can’t even settle a question that requires them to walk down the hallway.

It’s not even like we’re trying to detect a subtle effect here. Both sides agree that the signal is very large. They just disagree what direction it’s very large in!

A recent theme of this blog has been that Pyramid Of Scientific Evidence be damned, our randomized controlled trials suck so hard that a lot of the time we’ll get more trustworthy information from just looking at the ecological picture. Williams and Ceci have done this (see Part V, Section b of their supplement, “Do These Results Differ From Actual Hiring Data”) and report that studies of real-world hiring data confirm women have an advantage over men in STEM faculty hiring (although far fewer of them apply). It also matches the anecdotal evidence I hear from people in the field. I’m not necessarily saying I’m ambivalent between the two studies’ conclusions. Just that it bothers me that we have to go to tiebreakers after doing two good randomized controlled trials.

At this point, I think the most responsible thing would be to have a joint study by both teams, where they all agree on a fair protocol beforehand and see what happens. Outside of parapsychology I’ve never heard of people taking such a drastic step – who would get to be first author?! – but at this point it’s hard to deny that it’s necessary.

In conclusion, I believe the Moss-Racusin et al study more, but I think the Williams and Ceci study is more believable. And the best way to fight sexism in science is to remind people that it would be hard for women to make things any more screwed up than they already are.

# Polemical Imbalance

Today is an exciting day for me. I got argued against on Mad In America. This one is going straight to my resume.

Mad In America apparently doesn’t like being called an anti-psychiatry blog, so let’s call it a blog…that discusses psychiatry…and doesn’t usually like what it sees. They were heavily involved in popularizing the idea that psychiatry erred grieviously in overselling “chemical imbalance”, and they didn’t much like my post on the same topic:

Alexander argues that the notion that psychiatrists once promoted the idea of low serotonin as a cause of depression and Selective Serotonin Reuptake Inhibitors (SSRIs) as proper treatment for that deficiency is all simply a false “narrative” invented by “antipsychiatry” activists. These activists then “frame it as ‘proof’ that psychiatrists are drug company shills who were deceiving the public.” Alexander points to quotes of American Psychiatric Association officials in a post by MIA Blogger Philip Hickey, and notes that none of the quotes specifically describe a low-serotonin explanation for depression. The Hickey post cited is not actually about that topic, but about the promotion of the phrase “chemical imbalance”; nevertheless, Alexander broadly refers to Hickey and all of Mad in America as “antipsychiatry”, and he then writes, “If the antipsychiatry community had quotes of APA officials saying it’s all serotonin deficiency, don’t you think they would have used them?” Alexander argues, “The idea that depression is a drop-dead simple serotonin deficiency was never taken seriously by mainstream psychiatry.” There seems to be a lot of evidence to the contrary still today readily available even on the web, though.

This is exactly the sort of fight I probably shouldn’t get involved in continuing. But I’m going to do so anyway, because I think Mad In America’s counterargument is actually going to end up supporting my point and maybe shed more light on the situation.

Up there, when they say “Alexander points to quotes of American Psychiatric Association officials in a post by MIA Blogger Philip Hickey, and notes that none of the quotes specifically describe a low-serotonin explanation for depression [but] the Hickey post cited is not actually about that topic, but about the promotion of the phrase ‘chemical imbalance'” – that’s where I get pretty confident they’ve missed my point.

Remember, the thesis of my last post was that the “chemical imbalance” argument hides a sort of bait-and-switch going on between the following two statements:

(A): Depression is complicated, but it seems to involve disruptions to the levels of brain chemicals in some important way

(B): We understand depression perfectly now, it’s just a deficiency of serotonin.

If you equivocate between them, you can prove that psychiatrists were saying (A), and you can prove that (B) is false and stupid, and then it’s sort of like psychiatrists were saying something false and stupid.

Given that this is my thesis, it’s exactly right for me to debate a post on “chemical imbalance” by showing that none of the quotes involved reduce the problem to just a basic serotonin deficiency!

And when Rob Wipond from MIA says he’s found “a lot of evidence to the contrary still readily available even on the web”, well, spoiler, he’s found more people saying A.

II.

Let’s go through his examples:

For example, a 2004 Washington University in St. Louis press release, about a study published in Biological Psychiatry, states that the “brain’s serotonin receptors” are “at abnormally low levels in depressed people” and that antidepressants “work by increasing serotonin levels in the brain.”

I assume he’s talking about this press release about a study that shows abnormally low levels of serotonin receptors in depressed people. First of all, the study actually did show this. I don’t think it’s irresponsible to mention that a study shows low levels of serotonin receptors in depressed people when a study actually shows this. Second of all, the press release makes it extremely clear that they don’t know exactly what’s going on: “Little is understood about how depression makes people feel sad, but neuroscientists do know that the brain chemical serotonin is involved.” They mention that SSRIs appear to work for depression, but admit that “The bad news is that beyond that first step of increasing serotonin, we understand very little about how these drugs relieve symptoms of depression”. Finally, this study actually found something much more complicated than the prevailing narrative – a serotonin deficiency model of depression would have predicted high levels of serotonin receptors in related brain structures (more chemicals = fewer receptors) but in fact it found the opposite. This fits with the emerging theory that depression may be related to increased serotonin levels in certain parts of the brain, which SSRIs provoke a compensatory response against.

This press release is actually as good as the harshest critic could have wished for. It admits we don’t really know how depression works, it admits we don’t really know how SSRIs treat it, and then it presents the result of a study that shows that serotonin is implicated but not in the way the “serotonin deficiency” theory would expect.

The only way Mad In America turned this into a poster child for psychiatry deceiving people about serotonin was to quote from it extremely out of context.

Let’s go to their next example:

And there is prominent psychiatrist Richard Friedman writing in the New York Times in 2007 that psychiatrists were soon going to be able to conduct “a simple blood test” to determine “what biological type of depression” a person had and then treat them with the right drug. “For example,” writes Friedman, “some depressed patients who have abnormally low levels of serotonin respond to S.S.R.I.’s, which relieve depression, in part, by flooding the brain with serotonin.”

Okay, but Friedman starts with a story about how SSRIs often don’t work for patients, then says that this is because some people have depression that doesn’t seem to be serotonergic: “Some depressed patients who have abnormally low levels of serotonin respond to SSRIs, which relieve depression, in part, by flooding the brain with serotonin. Other depressed patients may have an abnormality in other neurotransmitters that regulate mood, like norepinephrine or dopamine, and may not respond to SSRIs”. He says (correctly!) that “in everyday clinical practice, we have little ability to predict what specific treatment will work for you”.

These are not the words of a drug company shill who says that depression is 100% serotonin in order to put everyone on SSRIs! These are the words of someone who agrees with me that depression is somehow related to neurotransmitters, but it’s still very uncertain which ones and how. His only sin seems to be an overly optimistic view of the speed at which we would come out with genetic tests.

Next example:

There’s also a lot of evidence that the low-serotonin theory of depression is still today being taken seriously by mainstream psychiatry and is still being promoted to the public. A current University of Bristol public education website on depression explains that, “Low serotonin levels are believed to be the cause of many cases of mild to severe depression.”

That appears to be this University of Bristol public education website. The site says it’s by “Claire Rosling”, so I searched her name and I get this roster of people’s sophomore chemistry projects. Ms. Rosling’s is…the website Mad In America cited. Apparently this was part of some college chemistry assignment where people write about molecules to compete for a £50 prize. Ms. Rosling’s was serotonin.

So Mad In America argues that the entire psychiatric establishment is pushing the “depression = serotonin” argument, but the best example they can come up with is some poor woman’s undergraduate chemistry homework?

(in case you’re wondering, she didn’t win. Some girl named Anna won for her webpage on Recycling Plastic.)

Next example!

A current Harvard Medical School special health report, “Understanding Depression”, explains that, “Research supports the idea that some depressed people have reduced serotonin transmission. Low levels of a serotonin byproduct have been linked to a higher risk for suicide.”

Once again, holy !@#$, they’re reporting the results of actual studies. It’s dishonest to do studies on serotonin and find that it is linked to depression? Anyway, when I look up the actual report it starts with the following paragraph: “It’s often said that depression results from a chemical imbalance, but that figure of speech doesn’t capture how complex the disease is. Research suggests that depression doesn’t spring from simply having too much or too little of certain brain chemicals. Rather, depression has many possible causes, including faulty mood regulation by the brain, genetic vulnerability, stressful life events, medications, and medical problems. It’s believed that several of these forces interact to bring on depression.”

Once again, this is the best you can do to find psychiatrists pushing an oversimplified version of the chemical imbalance theory??!

Next example:

WebMD’s “Depression Center” states that, “There are many researchers who believe that an imbalance in serotonin levels may influence mood in a way that leads to depression. Possible problems include low brain cell production of serotonin, a lack of receptor sites able to receive the serotonin that is made… According to Princeton neuroscientist Barry Jacobs… common antidepressant medications known as SSRIs, which are designed to boost serotonin levels, help kick off the production of new brain cells, which in turn allows the depression to lift.”

First of all, this page does not use the classic “serotonin deficiency” theory of depression. This is the hippocampal neurogenesis theory, which in my last post I specifically contrasted with the classic serotonin deficiency theory. Yes, it involves serotonin in some way, but since one of the most important facts about depression is that SSRIs treat it, every theory is going to involve serotonin in some way.

Further, right after this paragraph, WebMD continues: “Although it is widely believed that a serotonin deficiency plays a role in depression, there is no way to measure its levels in the living brain. Therefore, there have not been any studies proving that brain levels of this or any neurotransmitter are in short supply when depression or any mental illness develops. Blood levels of serotonin are measurable — and have been shown to be lower in people who suffer from depression – but researchers don’t know if blood levels reflect the brain’s level of serotonin. Also, researchers don’t know whether the dip in serotonin causes the depression, or the depression causes serotonin levels to drop.”

Once again, I see nothing here to indicate that they are covering up flaws in this theory, pushing it to unsuspecting consumers, or claiming that exploratory research is settled science. They’re presenting the best theories we’ve got, then noting how tentative they are and what the flaws are.

(on the other hand, the article does say that there are “40 million” brain cells, when in fact there are about 90 billion. I’m not saying you should trust WebMD, just that they don’t bungle depression in that particular way)

Next example:

And if the theory was never taken seriously and isn’t being taken seriously, no one has apparently told the National Academy of Sciences or two news media outlets with expert psychiatric editorial boards yet. Psychiatry Advisor’s February 12, 2015 headline for a report about a Duke University study is, “Serotonin Deficiency May Up Depression Risk.” Psychiatry Advisor explains that, “(m)ice with normal serotonin levels, the control group, did not demonstrate depression symptoms a week after the social stress, while the serotonin-deficient rodents did(.)” The study, appearing in the Proceedings of the National Academy of Sciences, states that, serotonin deficiency has been “implicated in the etiology of depression” though a cause-effect relationship has not yet been “formally established.” The researchers write that their results, “provide additional insight into the serotonin deficiency hypothesis of depression.” Medical News Today headline their report on it even more strongly: “Mouse study finds that serotonin deficiency does increase depression risk.” (Medical News Today notes in passing that an earlier, somewhat similar study by a different team came to the exact opposite findings.)

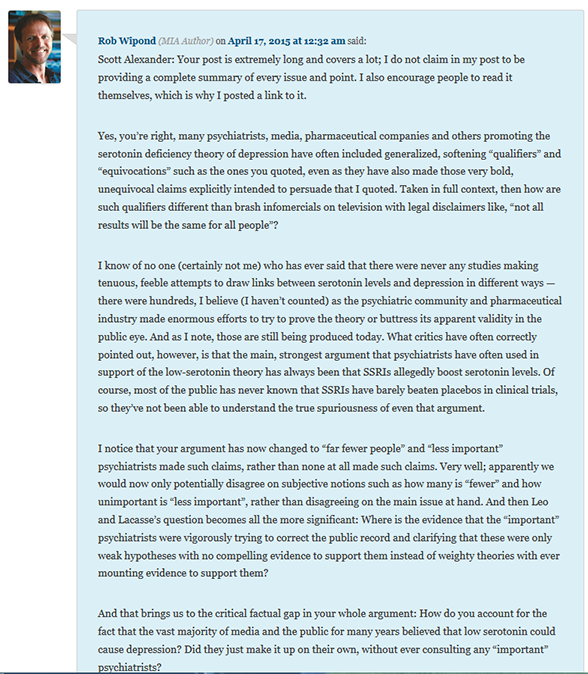
At this point Mad in America’s examples are self-refuting. I am getting the impression they will never be happy unless no news media ever covers the dozens of studies that come out each year linking depression to serotonin. I know this sounds mean, but what other conclusion am I supposed to come to? Here we have a study that provides some evidence for serotonin’s involvement, says very specifically that “a cause-effect relationship has not been formally established”, mentions that other studies have shown the opposite – and yet Mad In America still wants me to accept this as an example of irresponsibly pushing the serotonin theory!

Look. Hundreds of studies have shown some sort of relationship between serotonin and depression. At this point that’s not controversial. What’s controversial is the importance of the relationship, whether it’s causal, whether other things matter more, et cetera. Every single one of Mad In America’s examples has been pretty exemplary in saying that all of these things are still uncertain and need to be investigated further. What more could they do to be more responsible? A total blackout on all news coverage of the new evidence for serotonin’s involvement that keeps coming in?

Ironically, if people had done that, we would have far less evidence that depression was not just a simple serotonin deficiency. The most important nail in that theory’s coffin was that tianeptine, a medication that lowers serotonin levels, effectively treats depression. But that’s a study about serotonin of exactly the same sort as the University of Washington study Mad In America complains about! One of the most convincing alternatives to a purely serotonergic picture is the BDNF-neurogenesis theory. But that’s exactly the theory being pushed in the WebMD article Mad In America complains about!

III.

I raised some of these issues in a comment on the Mad in America blog, and author Rob Wipond kindly responded to me:



Let me address some of these objections piece by piece:

Yes, you’re right, many psychiatrists, media, pharmaceutical companies and others promoting the serotonin deficiency theory of depression have often included generalized, softening “qualifiers” and “equivocations” such as the ones you quoted, even as they have also made those very bold, unequivocal claims explicitly intended to persuade that I quoted. Taken in full context, then how are such qualifiers different than brash infomercials on television with legal disclaimers like, “not all results will be the same for all people”?

The very bold, unequivocal claims explicitly intended to persuade that you quoted WERE A SOPHOMORE CHEMISTRY PROJECT, PLUS A BUNCH OF PEOPLE SPECIFICALLY SAYING THAT THESE SHOULD NOT BE TAKEN AS VERY BOLD UNEQUIVOCAL CLAIMS, BUT THEN YOU QUOTED OUT OF CONTEXT TO TAKE THAT PART OUT. AND THE SOPHOMORE CHEMISTRY PROJECT DIDN’T EVEN WIN THE £50 PRIZE.

Okay. Sorry. I shouldn’t have yelled like that. More seriously: there are a lot of things we don’t totally understand, but which scientific research suggests some weak preliminary theories about. For example, we don’t understand fibromyalgia, but if I were writing a textbook on fibromyalgia, or if a patient asked me what it was, then after some appropriate caveats and equivocations, I would say it has something to do with some sort of inflammation in the fascia which causes central sensitization to pain stimuli. Could I end up being totally wrong? Yeah. But at this point I think there’s enough evidence in this direction that, insofar as it’s important to satisfy patients’ curiosity about what’s going on with them, that it’s proper to mention the current best guess. Likewise, if I am a researcher or a scientific publication, I don’t think I have some duty to carefully hide my results. A big part of scientific progress is people saying “I just got some small amount of evidence which makes me think it’s this” and then other people trying to confirm or refute that with more evidence, until eventually it comes together into a strong theory.

I think researchers and psychiatrists were pretty responsible in coming up with the serotonin deficiency theory. It was inspired by the effectiveness of serotonergic drugs. Then a bunch of studies – Wipond agrees there were hundreds – provided results that seemed to confirm it. Given all of this information, I don’t think it was negligent to say that there was quite a bit of evidence pointing to serotonin, as long as you followed this with caveats that the evidence was still preliminary and lots of other things seemed to be involved too. As I’ve been arguing all along, that’s exactly what most people did.

I know of no one (certainly not me) who has ever said that there were never any studies making tenuous, feeble attempts to draw links between serotonin levels and depression in different ways — there were hundreds, I believe (I haven’t counted) as the psychiatric community and pharmaceutical industry made enormous efforts to try to prove the theory or buttress its apparent validity in the public eye. And as I note, those are still being produced today. What critics have often correctly pointed out, however, is that the main, strongest argument that psychiatrists have often used in support of the low-serotonin theory has always been that SSRIs allegedly boost serotonin levels. Of course, most of the public has never known that SSRIs have barely beaten placebos in clinical trials, so they’ve not been able to understand the true spuriousness of even that argument.

I don’t really understand this objection. There was a strong piece of evidence in favor of serotonin in the form of SSRI-effectiveness, scientists pursued that lead by doing hundreds of studies implicating serotonin using different methodologies, most were in favor and so scientists thought the theory had some merit…what exactly is wrong here? This sounds like every scientific theory – Wegener noted that continents looked like they fit together in a way that implied continental drift, geologists did hundreds of other studies that all pointed to continental drift, therefore they started believing in continental drift.

While Wipond may not know of the people saying there was no evidence for serotonin besides SSRI effectiveness, these people certainly exist and provide one of his side’s major arguments. Indeed, many of the articles I linked to on my original post made exactly that argument. The BBC said that: “although ideas like the serotonin theory of depression have been widely publicised, scientific research has not detected any reliable abnormalities of the serotonin system in people who are depressed.” New York Review Of Books says: “Instead of developing a drug to treat an abnormality, an abnormality was postulated to fit a drug…But the main problem with the theory is that after decades of trying to prove it, researchers have still come up empty-handed”.

To learn more about the claim that SSRIs barely beat placebo, see my article on this.

I notice that your argument has now changed to “far fewer people” and “less important” psychiatrists made such claims, rather than none at all made such claims. Very well; apparently we would now only potentially disagree on subjective notions such as how many is “fewer” and how unimportant is “less important”, rather than disagreeing on the main issue at hand. And then Leo and Lacasse’s question becomes all the more significant: Where is the evidence that the “important” psychiatrists were vigorously trying to correct the public record and clarifying that these were only weak hypotheses with no compelling evidence to support them instead of weighty theories with ever mounting evidence to support them?

In retrospect, “no one has ever said” is a stupid thing for me to have said. I do not deny that a sophomore at University of Bristol once said low serotonin caused depression. And you can find individual psychiatrists who believe a lot of stupid stuff. Some psychiatrists believe in homeopathy. Some psychiatrists believe in reincarnation (the guy in that article conducted my job interview at the University of Virginia. I tried to be very polite.) Some psychiatrists believe that after losing hundreds of thousands of dollars to online Nigerian scammers, it makes perfect sense to give hundreds of thousands more dollars to other Nigerian scammers, because “these were different Nigerians”. But I will venture to say none of these are consensus positions in the psychiatric community.

And this is why I wanted to continue this discussion here on this blog. If I had selected a set of statements from eminent psychiatrists that had lots of caveats and were extremely responsible, I could be justifiably accused of cherry-picking. Instead, Mad In America selected some statements, probably intending to cherry-pick the other way, but when looked at more closely, they’re all pretty responsible and say exactly what I would have said at the time – SSRIs seem to work, there’s some evidence pointing to serotonin being involved, but the whole thing is terribly complicated. To me, this establishes the consensus position in a way much more clearly than I could have done on my own.

And yes, this consensus position got simplified and distorted. I have no doubt that drug companies drew from it to do exactly the sort of infomercials that Mr. Wipond describes. I have no doubt that individual psychiatrists, when faced with low-functioning patients who are bad at understanding complicated systems but who really wanted to know what was going on, said “serotonin” and left it at that. And I have no doubt that to a public who still largely think evolution means “once upon a time a chimp gave birth to a human baby”, complicated caveats about how serotonin levels are linked to depression but might not cause depression largely went over their heads except for the single word “serotonin”.

(“That’s the happiness molecule! Right?”)

But in general I think my point stands. “Chemical imbalance” as generally used points to a sophisticated model of interacting metabolic pathways which goes far beyond serotonin, and which as far as I know is still very much on the table. While serotonin was justifiably pointed to as a promising candidate early on, it was generally done with appropriate caveats that turned out to be warranted, and the research community has now retreated from some of that earlier language while still considering serotonin a promising lead. And SSRIs continue to be moderately effective antidepressants in the people for whom they are indicated.

(ie somewhere less than half of the people for whom they are prescribed).

# Blame Theory

It’s always dangerous to speculate about the hidden psychological motives of people you disagree with – this is the sin of Bulverism. But like most sins, it’s also fun. So please forgive me while I talk about blame.

Many people have remarked on the paradox of an academia made mostly of upper-class ethnic-majority Westerners trying so very hard to find reasons why lots of things are the fault of upper-class ethnic-majority Westerners. The simplest example I can think of is attributing the woes of Third World countries to colonialism; without meaning to trivialize the evils of colonization, a lot of academics seem to go beyond what even the undeniably awful facts can support. Dependency theory, for example, is now mostly discredited, as are a lot of the Marxist perspectives. I would provide other examples if I weren’t satisfied you can generate them independently.

This is on the face of it surprising; naively we would expect people to cast themselves and those like them in as positive a light as possible. Forget about whether these attributions of blame are right or wrong. Even if they were right I would not expect people to believe them as enthusiastically as they do.

The theories I’ve heard to explain this paradox are rarely very flattering; usually something about class signaling, or holier-than-thou-ness, or trying to justify the existence of an academic elite.

I want to propose another possibility: what if people are really, fundamentally, good?

Moral philosophy distinguishes between a couple of ethical systems, like deontology, utilitarianism and virtue ethics. Most people without philosophical training settle into a sort of mishmash of all of them, but one which, I think, is closer to deontology than either of the others. Call it Moral Therapeutic Deontology. Like all deontological systems, it focuses on following certain rules: don’t murder, don’t steal, respect your parents, pay back your debts. Like all deontological systems, other things like charity are “supererogatory”, meaning they’re nice but not really necessary. If you’ve got extra time and energy after doing the important stuff, then sure, do the superogatory stuff, whatever, but it’s hardly where your moral focus should be.

On the other hand, when confronted with the full extent of human suffering – whether by living in a poor area, or serving in a war zone, or traveling to a Third World country, or treating depression patients – it’s hard to think about anything else. The sheer burning horribleness of it becomes this unscratchable itch, this flaw in the world that blots out the sun.

And here’s Moral Therapeutic Deontology, saying, “Yeah, helping quench the burning fire of human suffering is nice, but it’s not like a real thing that real morality should care about. It’s not your duty.”

This is some heavy cognitive dissonance. It doesn’t match basic intuitions about the importance of the matter. Even worse, it doesn’t allow you to communicate the importance of the matter to other people. If you say “Look at all these people living squalid and miserable in the slums without any hope,” and they say “Yeah, well, it would be supererogatory to help them and I’m not feeling supererogatory today,” you don’t really have a leg to stand on.

There’s an easy way to resolve the dissonance without abandoning either Moral Therapeutic Deontology or your concern for the less well-off. That resolution is to prove that human suffering is you and your friends’ fault. Deontology very clearly says that if you cause a problem, it’s your job to help fix it. If you can prove that the reason the Third World is suffering is because of First World white people, you have a strong claim that you as a First World white person should be deeply emotionally invested in solving it; that your friends and neighbors, as First World white people, ought to help you; and that your government, as that of a First World majority-white country, is justified in using taxpayer money to get involved.

I think this might be a part of what’s happening. People feel a need to help the less-advantaged so strongly that they come up with a justification to do so that makes sense in their own moral system, whether it’s factually accurate or not.

I am not as fanatical a partisan of utilitarianism as I used to be, but this still seems like one of the situations where it has an obvious advantage. Utilitarianism tells us that we are perfectly justified in seeing the relief of suffering as a pressing need. We don’t need to justify it by positing facts that may later be proven untrue; it is self-justifying. People sometimes complain that a flaw of utilitarianism is that it implies a heavy moral obligations to help all kinds of people whether or not any of their problems are our fault; the world is divided between those who consider that a bug and those who find it a very helpful feature.

I want more people to become familiar with utilitarianism because I think a lot of the colonialism theory stuff is net hurtful. It combines a justification for helping the poor with an insult to people’s identity, and probably makes the former less palatable to many people than it would be naturally. It also makes our need to help the poor hinge on an empirical point; if that empirical point gets disproved, things become pretty awkward.

This theory implies that utilitarian liberals will have all the features of liberalism except the interest in blaming their own group for major problems. My anecdotal experience confirms that. The utilitarians I know are very interested in helping the poor and in various other liberal ideas, but are more likely than other liberals to roll their eyes at talk about colonialism and stereotype threat. I think it’s because they feel confident in their right to care about the disadvantaged regardless.

# Universal Love, Said The Cactus Person

“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, 258,952,897, 654,000,350, 692,006,139?

“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, 258,952,897, 654,000,350, 692,006,139?

“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.

“Or maybe you guys are so intoxicated on spiritual wisdom that you couldn’t think straight if your life depended on it. Maybe 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. Maybe 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, but 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 daath, 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.”

# Growth Mindset 3: A Pox On Growth Your Houses

[EDIT: The author of this paper has responded; I list his response here.]

Jacques Derrida proposed a form of philosophical literary criticism called deconstruction. I’ll be the first to admit I don’t really understand it, but it seems to have something to do with assuming all texts secretly contradict their stated premise and apparent narrative, then hunting down and exposing the plastered-over areas where the author tries to hide this.

I have no idea whether this works for literature or not, but it’s a useful way to read scientific papers.

Consider a popular field – or, at least, a field where a certain position is popular. For example, we’ve been talking a lot about growth mindset recently. There seem to be a lot of researchers working to prove growth mindset and not a lot working to disprove it. Journals are pretty interested in studies showing growth mindset interventions work, and maybe not so interested in studies showing they don’t. I’ll admit that my strong suspicions of publication bias don’t seem to be borne out by the facts here – see this meta-analysis – but I bet its more sinister cousin “all experimenters believe the same thing and have the same experimenter effects” bias is alive and well.

In a field like that, you’re not going to get the contrarian studies you want, but one way to find the other side of the issue is to look a little more closely at the studies that do get published, the ones that say they’re in support of the thesis, and see if you can find anything incriminating.

Here’s a perfect example: Mindset Interventions Are A Scalable Treatment For Academic Underachievement, by a team of six researchers including Carol Dweck.

The abstract reads:

The efficacy of academic-mind-set interventions has been demonstrated by small-scale, proof-of-concept interventions, generally delivered in person in one school at a time. Whether this approach could be a practical way to raise school achievement on a large scale remains unknown. We therefore delivered brief growth-mind-set and sense-of-purpose interventions through online modules to 1,594 students in 13 geographically diverse high schools. Both interventions were intended to help students persist when they experienced academic difficulty; thus, both were predicted to be most beneficial for poorly performing students. This was the case. Among students at risk of dropping out of high school (one third of the sample), each intervention raised students’ semester grade point averages in core academic courses and increased the rate at which students performed satisfactorily in core courses by 6.4 percentage points. We discuss implications for the pipeline from theory to practice and for education reform.

This sounds really, really impressive! It’s hard to imagine any stronger evidence in growth mindset’s favor.

And then you make the mistake of reading the actual paper.

The paper asked a 1,594 students from a bunch of different high schools to take a 45 minute online course.

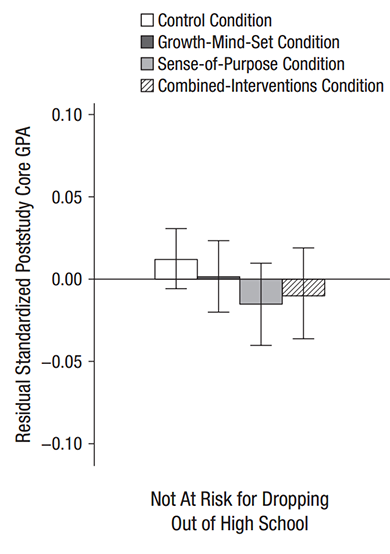
A quarter of the students took a placebo course that just presented some science about how different parts of the brain do different stuff.

Another quarter took a course that was supposed to teach growth mindset.

Still another quarter took a course about “sense of purpose” which talked about how schoolwork was meaningful and would help them accomplish lots of goals and they should be happy to do it. This was also classified as a “mindset intervention”, though it seems pretty different.

And the final quarter took both the growth mindset course and the “sense of purpose” course.

Then they let all students continue taking their classes for the rest of the semester and saw what happened, which was this:



[EDIT: I totally bungled these graphs! See discussion of exactly how on the author’s reply above, without which the information below will be misleading at best]

Among ordinary students, the effect on the growth mindset group was completely indistinguishable from zero, and in fact they did nonsignificantly worse than the control group. This was the most basic test they performed, and it should have been the headline of the study. The study should have been titled “Growth Mindset Intervention Totally Fails To Affect GPA In Any Way”.

Instead they went to subgroup analysis. Subgroup analysis can be useful to find more specific patterns in the data, but if it’s done post hoc it can lead to what I previously called the Elderly Hispanic Woman Effect, after medical papers that can’t find their drug has any effect on people at large, so they keep checking different subgroups – young white men…nothing. Old black men…nothing. Middle-aged Asian transgender people…nothing. Newborn Australian aboriginal butch lesbians…nothing. Elderly Hispanic women…p = 0.049…aha! And the study gets billed as “Scientists Find Exciting New Drug That Treats Diabetes In Elderly Hispanic Women.”

As per the abstract, the researchers decided to focus on an “at risk” subgroup because they had principled reasons to believe mindset interventions would work better on them. In their subgroup of 519 students who had a GPA of 2.0 or less last semester, or who failed one or more academic courses last semester:



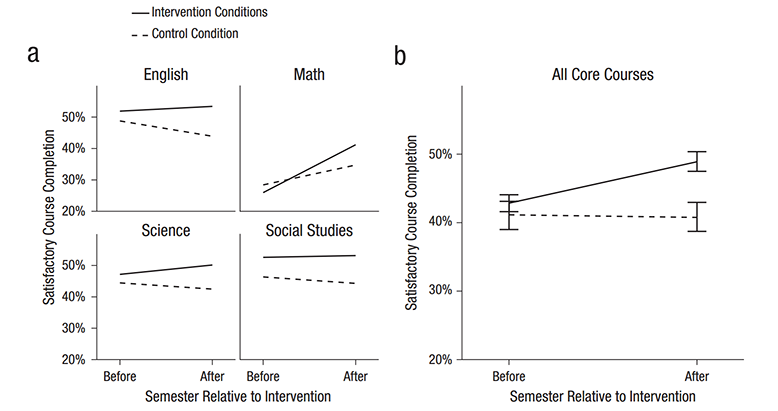
Growth mindset still doesn’t differ from zero. And growth mindset does nonsignificantly worse than their “sense of purpose” intervention where they tell children to love school. In fact, the students who take both “sense of purpose” and growth mindset actually do (nonsignificantly) worse than sense-of-purpose alone!

But the control group mysteriously started doing much worse in all their classes right after the study started, so growth mindset is significantly better than the control group. Hooray!

Why would the control group’s GPA suddenly decline? The simplest answer would be that by coincidence the class got harder right after the study started, and only the intervention kids were resilient enough to deal with it – but that can’t be right, because this was done at eleven different schools, and they wouldn’t have all had their coursework get harder at the same time.

Another possibility is that sufficiently low-functioning kids are always declining – that is, as time goes on they get more and more behind in their coursework, so their grades at time t+1 are always less than at time t, and maybe growth mindset has arrested this decline. This is plausible and I’d be interested in seeing if other studies have found this.

Perhaps aware that this is not very convincing, the authors go on to do another analysis, this one of percent of students passing their classes.



This is the same group of at-risk students as the last one. It’s graphing what percent of these students pass versus fail their courses. The graph on th left shows that a significantly higher number of students in the intervention conditions pass their courses than in the control condition.

This is better, but one part still concerns me.

Did you catch that phrase “intervention conditions”? The authors of the study write: “Because our primary research question concerned the efficacy of academic mindset interventions in general when delivered via online modules, we then collapsed the intervention conditions into a single intervention dummy code (0 = control, 1 = intervention).

We don’t know whether growth mindset did anything for even these students in this little subgroup, because it was collapsed together with the (more effective) “sense of purpose” intervention before any of these tests were done. I don’t know if this is just for convenience, or if it is to obfuscate that it didn’t work on its own.

[EDIT: Scott McGreal looks further and finds in the supplementary material that growth mindset alone did NOT significantly improve pass rates!]

The abstract of this study tells you none of this. It just says: “Mindset Interventions Are A Scalable Treatment For Academic Overachievement…Among students at risk of dropping out of high school (one third of the sample), each intervention raised students’ semester grade point averages in core academic courses and increased the rate at which students performed satisfactorily in core courses by 6.4 percentage points” From the abstract, this study is a triumph.

But my own summary of these results, as relevant to growth mindset is as follows:

For students with above a 2.0 GPA, a growth mindset intervention did nothing.

For students with below a 2.0 GPA, the growth mindset interventions may not have improved GPA, but may have prevented GPA from falling, which for some reason it was otherwise going to do.

Even in those students, it didn’t do any better than a “sense-of-purpose” intervention where children were told platitudes about how doing well in school will “make their families proud” and “make a positive impact”.

In no group of students did it significantly increase chance of passing any classes.

Haishan writes:

If ye read only the headlines, what reward have ye? Do not even the policymakers the same? And if ye take the abstract at its face, what do ye more than others? Do not even the science journalists so?”

Titles, abstracts, and media presentations are where authors can decide how to report a bunch of different, often contradictory results in a way that makes it look like they have completely proven their point. A careful look at the study may find that their emphasis is misplaced, and give you more than enough ammunition against a theory even where the stated results are glowingly positive.

The only reason we were told these results is that they were in the same place as a “sense of purpose mindset” intervention that looked a little better, so it was possible to publish the study and claim it as a victory for mindsets in general. How many studies that show similar results for growth mindset lack a similar way of spinning the data, and so never get seen at all?

# Nefarious Nefazodone And Flashy Rare Side Effects

[Epistemic status: I am still in training. I am not an expert on drugs. This is poorly-informed speculation about drugs and it should not be taken seriously without further research. Nothing in this post is medical advice.]

I.

Which is worse – ruining ten million people’s sex lives for one year, or making one hundred people’s livers explode?

I admit I sometimes use this blog to speculate about silly moral dilemmas for no reason, but that’s not what’s happening here. This is a real question that I deal with on a daily basis.

SSRIs, the class which includes most currently used antidepressants, are very safe in the traditional sense of “unlikely to kill you”. Suicidal people take massive overdoses of SSRIs all the time, and usually end up with little more than a stomachache for their troubles. On the other hand, there’s increasing awareness of very common side effects which, while not disabling, can be pretty unpleasant. About 50% of users report decreased sexual abilities, sometimes to the point of total loss of libido or anorgasmia. And something like 25% of users experience “emotional blunting” and the loss of ability to feel feelings normally.

Nefazodone (brand name Serzone®, which would also be a good brand name for a BDSM nightclub) is an equally good (and maybe better) antidepressant that does not have these side effects. On the other hand, every year, one in every 300,000 people using nefazodone will go into “fulminant hepatic failure”, which means their liver suddenly and spectacularly stops working and they need a liver transplant or else they die.

There are a lot of drug rating sites, but the biggest is Drugs.com. 467 Drugs.com users have given Celexa, a very typical SSRI, an average rating of 7.8/10. 14 users have given nefazodone an average rating of 9.1/10.

CrazyMeds might not be as dignified as Drugs.com, but they have a big and well-educated user base and they’re psych-specific. Their numbers are 3.3/5 (n = 253) for Celexa and 4.1/5 (n = 47) for nefazodone.

So both sites’ users seem to agree that nefazodone is notably better than Celexa, in terms of a combined measure of effectiveness and side effects.

But nefazodone is practically never used. It’s actually illegal in most countries. In the United States, parent company Bristol-Myers Squibb (which differs from normal Bristol-Myers in that it was born without innate magical ability) withdrew it from the market, and the only way you can find it nowadays is to get it is from an Israeli company that grabbed the molecule after it went off-patent. In several years working in psychiatry, I have never seen a patient on nefazodone, although I’m sure they exist somewhere. I would estimate its prescription numbers are about 1% of Celexa’s, if that.

The problem is the hepatic side effects. Nobody wants to have their liver explode.

But. There are something like thirty million people in the US on antidepressants. If we put them all on nefazodone, that’s about a hundred cooked livers per year. If we put them all on SSRIs, at least ten million of them will get sexual side effects, plus some emotional blunting.

My life vastly improved when I learned there was a searchable database of QALYs for different conditions. It doesn’t have SSRI-induced sexual dysfunction, but it does have sexual dysfunction due to prostate cancer treatment, and I assume that sexual dysfunction is about equally bad regardless of what causes it. Their sexual dysfunction has some QALY weights averaging about 0.85. Hm.

Assume everyone with fulminant liver failure dies. That’s not true; some get liver transplants, maybe some even get a miracle and recover. But assume everyone dies – and further, they die at age 30, cutting their lives short by fifty years.

In that case, putting all depressed people on nefazodone for a year costs 5,000 QALYs, but putting all depressed people on SSRIs for a year costs 1,500,000 QALYs. The liver failures may be flashier, but the 3^^^3 dust specks worth of poor sex lives add up to more disutility in the end.

I don’t want to overemphasize this particular calculation for a couple of reasons. First, SSRIs and nefazodone both have other side effects besides the major ones I’ve focused on here. Second, I don’t know if the level of SSRI-induced sexual dysfunction is as bad as the prostate-surgery-induced sexual dysfunction on the database. Third, there are a whole bunch of antidepressants that are neither SSRIs nor nefazodone and which might be safer than either.

But I do want to emphasize this pattern, because it recurs again and again.

II.

In that spirit, which would you rather have – something like a million people addicted to amphetamines, or something like ten people have their skin eat itself from the inside?

I can’t get good numbers on how many adults abuse Adderall, but a quick glance at the roster for my hospital’s rehab unit suggests “a lot”. Huffington Post calls it the most abused prescription drug in America, which sounds about right to me. Honestly there are worse things to be addicted to than Adderall, but it’s not completely without side effects. The obvious ones are anxiety, irritability, occasionally frank psychosis, and sometimes heart problems – but a lot of the doctors I work with go beyond what the research can really prove and suggest it can produce lasting negative personality change and predispose people to other forms of addictive and impulsive behavior.

If you’ve got to give adults a stimulant, I would much prefer modafinil. It’s not addictive, it lacks most of Adderall’s side effects, and it works pretty well. I’ve known many people on modafinil and they give it pretty universally positive reviews.

On the other hand, modafinil may or may not cause a skin reaction called Stevens Johnson Syndrome/Toxic Epidermal Necrolysis, which like most things with both “toxic” and “necro” in the name is really really bad. The original data suggesting a connection came from kids, who get all sorts of weird drug effects that adults don’t, but since then some people have claimed to have found a connection with adults. Some people get SJS anyway just by bad luck, or because they’re taking other drugs, so it’s really hard to attribute cases specifically to modafinil.

Gwern’s Modafinil FAQ mentions an FDA publication which argues that the background rate of SJS/TEN is 1-2 per million people per year, but the modafinil rate is about 6 per million people per year. However, there are only three known cases of a person above age 18 on modafinil getting SJS/TEN, and this might not be different from background rates after all. Overall the evidence that modafinil increases the rate of SJS/TEN in adults at all is pretty thin, and if it does, it’s as rare as hen’s teeth (in fact, very close to the same rate as liver failure from nefazodone).

(also: consider that like half of Silicon Valley is on modafinil, yet San Francisco Bay is not yet running red with blood.)

(also: ibuprofen is linked to SJS/TEN, with about the same odds ratio as modafinil, but nobody cares, and they are correct not to care.)

I said I’ve never seen a doctor prescribe nefazodone in real life; I can’t say that about modafinil. I have seen one doctor prescribe modafinil. It happened like this: a doctor I was working with was very upset, because she had an elderly patient with very low energy for some reason, I can’t remember, maybe a stroke, and wanted to give him Adderall, but he had a heart arrythmia and Adderall probably wouldn’t be safe for him.

I asked “What about modafinil?”

She said, “Modafinil? Really? But doesn’t that sometimes cause Stevens Johnson Syndrome?”

And then I glared at her until she gave in and prescribed it.

But this is very, very typical. Doctors who give out Adderall like candy have no associations with modafinil except “that thing that sometimes causes Stevens-Johnson Syndrome” and are afraid to give it to people.

III.

Nefazodone and modafinil are far from the only examples of this pattern. MAOIs are like this too. So is clozapine. If I knew more about things other than psychiatry, I bet I could think of examples from other fields of medicine.

And partially this is natural and understandable. Doctors swear an oath to “first do no harm”, and toxic epidermal necrolysis is pretty much the epitome of harm. Thought experiments like torture vs dust specks suggest that most people’s moral intuitions say that no amount of aggregated lesser harms like sexual side effects and amphetamine addictions can equal the importance of avoiding even a tiny chance of some great harm like liver failure or SJS/TEN. Maybe your doctor, if you asked her directly, would endorse a principled stance of “I am happy to give any number of people anxiety and irritability in order to avoid even the smallest chance of one case of toxic epidermal necrolysis.”

And yet.

The same doctors who would never dare give nefazodone, consider Seroquel a perfectly acceptable second-line treatment for depression. Along with other atypical antipsychotics, Seroquel raises the risk of sudden cardiac death by about 50%. The normal risk of cardiac sudden death in young people is about 10 in 100,000 per year, so if my calculations are right, low-dose Seroquel causes an extra cardiac death once per every 20,000 patient-years. That’s ten times as often as nefazodone causes an extra liver death.

Yet nefazodone was taken off of the market by its creators and consigned to the dustbin of pharmacological history, and Seroquel is the sixth-best-selling drug in the United States, commonly given for depression, simple anxiety, and sometimes even to help people sleep.

Why the disconnect? Here’s a theory: sudden cardiac death happens all the time; sometimes God just has it in for you and your heart stops working and you die. Antipsychotics can increase the chances of that happening, but it’s a purely statistical increase, such that we can detect it aggregated over large groups but never be sure that it played a role in any particular case. The average person who dies of Seroquel never knows they died of Seroquel, but the average person who dies from nefazodone is easily identified as a nefazodone-related death. So nefazodone gets these big stories in the media about this young person who died by taking this exotic psychiatric drug, and it becomes a big deal and scares the heck out of everybody. When someone dies of Seroquel, it’s just an “oh, so sad, I guess his time has come.”

But the end result is this. When treatment with an SSRI fails, nefazodone and Seroquel naively seem to be equally good alternatives. Except nefazodone has a death rate of 1/300,000 patient years, and Seroquel 1/20,000 patient years. And yet everyone stays the hell away from the nefazodone because it’s known to be unsafe, and chooses the Seroquel.

I conclude either doctors are terrible at thinking about risk, or else maybe a little too good at thinking about risk.

I bring up the latter option because there’s a principal-agent problem going on here. Doctors want to do what’s best for their patients. But they also want to do what’s best for themselves, which means not getting sued. No one has ever sued their doctor because they got a sexual side effect from SSRIs, but if somebody dies because they’re the lucky 1/300,000 who gets liver failure from nefazodone, you can bet their family’s going to sue. Suddenly it’s not a matter of comparing QALYs, it’s a matter of comparing zero percent chance of lawsuit with non-zero percent chance of lawsuit.

(Fermi calculation: if a doctor has 100 patients at a time on antidepressants, and works for 30 years, then if she uses Serzone as her go-to antidepressant, she’s risking a 1% chance of getting the liver failure side effect once in her career. That’s small, but since a single bad lawsuit can bankrupt a doctor, it’s worth taking seriously.)

And that would be a tough lawsuit to fight. “Yes, Your Honor, I knew when I prescribed this drug that it sometimes makes people’s livers explode, but the alternative often gives people a bad sex life, and according to the theory of utilitarianism as propounded by 18th century philosopher Jeremy Bentham – ” … “Bailiff, club this man”.

And the same facet of nefazodone that makes it exciting for the media makes it exciting for lawsuits. When someone dies of nefazodone toxicity, everyone knows. When someone dies of Seroquel, “oh, so sad, I guess his time has come”.

That makes Seroquel a lot safer than nefazodone. Safer for the doctor, I mean. The important kind of safer.

This is why, as I mentioned before, I hate lawsuits as a de facto regulatory mechanism. Our de jure regulatory mechanism, the FDA, is pretty terrible, but to its credit it hasn’t banned nefazodone. One time it banned clozapine because of a flashy rare side effect, but everyone yelled at them and they apologized and changed their mind. With lawsuits there’s nobody to yell at, so we just end up with people very quietly adjusting their decisions in the shadows and nobody else being any the wiser.

I don’t want to overemphasize this, because I think it’s only one small part of the problem. After all, a lot of countries withdrew nefazodone entirely and didn’t even give lawsuits a chance to enter the picture.

But whatever the cause, the end result is that drugs with rare but spectacular side effects get consistently underprescribed relative to drugs with common but merely annoying side effects, or drugs that have more side effects but manage to hide them better.

# Prescriptions, Paradoxes, and Perversities

[WARNING: I am not a pharmacologist. I am not a researcher. I am not a statistician. This is not medical advice. This is really weird and you should not take it too seriously until it has been confirmed]

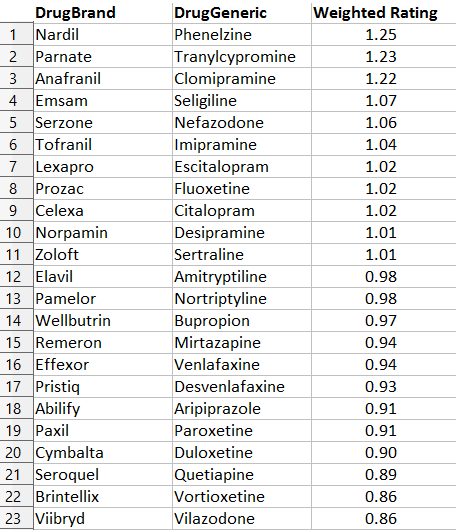
I.

I’ve been playing around with data from Internet databases that aggregate patient reviews of medications.

Are these any good? I looked at four of the largest such databases – Drugs.com, WebMD, AskAPatient, and DrugLib – as well as psychiatry-specific site CrazyMeds – and took their data on twenty-three major antidepressants. Then I correlated them with one another to see if the five sites mostly agreed.

Correlations between Drugs.com, AskAPatient, and WebMD were generally large and positive (around 0.7). Correlations between CrazyMeds and DrugLib were generally small or negative. In retrospect this makes sense, because these two sites didn’t allow separation of ratings by condition, so for example Seroquel-for-depression was being mixed with Seroquel-for-schizophrenia.

So I threw out the two offending sites and kept Drugs.com, AskAPatient, and WebMD. I normalized all the data, then took the weighted average of all three sites. From this huge sample (the least-reviewed drug had 35 ratings, the most-reviewed drug 4,797) I obtained a unified opinion of patients’ favorite and least favorite antidepressants.



This doesn’t surprise me at all. Everyone secretly knows Nardil and Parnate (the two commonly-used drugs in the MAOI class) are excellent antidepressants1. Oh, nobody will prescribe them, because of the dynamic discussed here, but in their hearts they know it’s true.

Likewise, I feel pretty good to see that Serzone, which I recently defended, is number five. I’ve had terrible luck with Viibryd, and it just seems to make people taking it more annoying, which is not a listed side effect but which I swear has happened.

The table also matches the evidence from chemistry – drugs with similar molecular structure get similar ratings, as do drugs with similar function. This is, I think, a good list.

Which is too bad, because it makes the next part that much more terrifying.

II.

There is a sixth major Internet database of drug ratings. It is called RateRx, and it differs from the other five in an important way: it solicits ratings from doctors, not patients. It’s a great idea – if you trust your doctor to tell you which drug is best, why not take advantage of wisdom-of-crowds and trust all the doctors?

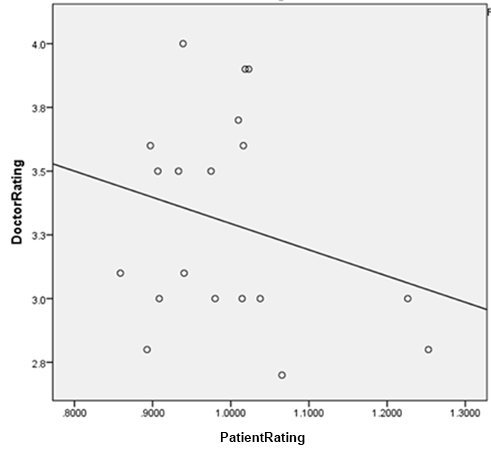


The RateRX logo. Spoiler: this is going to seem really ironic in about thirty seconds.

RateRx has a modest but respectable sample size – the drugs on my list got between 32 and 70 doctor reviews. There’s only one problem.

You remember patient reviews on the big three sites correlated about +0.7 with each other, right? So patients pretty much agree on which drugs are good and which are bad?

Doctor reviews on RateRx correlated at -0.21 with patient reviews. The negative relationship is nonsignificant, but that just means that at best, doctor reviews are totally uncorrelated with patient consensus.



This has an obvious but very disturbing corollary. I couldn’t get good numbers on how times each of the antidepressants on my list were prescribed, because the information I’ve seen only gives prescription numbers for a few top-selling drugs, plus we’ve got the same problem of not being able to distinguish depression prescriptions from anxiety prescriptions from psychosis prescriptions. But total number of online reviews makes a pretty good proxy. After all, the more patients are using a drug, the more are likely to review it.

Quick sanity check: the most reviewed drug on my list was Cymbalta. Cymbalta was also the best selling antidepressant of 2014. Although my list doesn’t exactly track the best-sellers, that seems to be a function of how long a drug has been out – a best-seller that came out last year might have only 1/10th the number of reviews as a best-seller that came out ten years ago. So number of reviews seems to be a decent correlate for amount a drug is used.

In that case, amount a drug is used correlates highly (+0.67, p = 0.005) with doctors’ opinion of the drug, which makes perfect sense since doctors are the ones prescribing it. But amount the drug gets used correlates negatively with patient rating of the drug (-0.34, p = ns), which of course is to be expected given the negative correlation between doctor opinion and patient opinion.

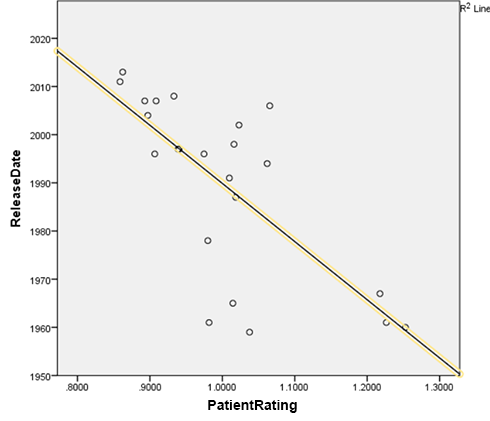
So the more patients like a drug, the less likely it is to be prescribed2.

III.

There’s one more act in this horror show.

Anyone familiar with these medications reading the table above has probably already noticed this one, but I figured I might as well make it official.

I correlated the average rating of each drug with the year it came on the market. The correlation was -0.71 (p < .001). That is, the newer a drug was, the less patients liked it3.



This pattern absolutely jumps out of the data. First- and second- place winners Nardil and Parnate came out in 1960 and 1961, respectively; I can’t find the exact year third-place winner Anafranil came out, but the first reference to its trade name I can find in the literature is from 1967, so I used that. In contrast, last-place winner Viibryd came out in 2011, second-to-last place winner Abilify got its depression indication in 2007, and third-to-last place winner Brintellix is as recent as 2013.

This result is robust to various different methods of analysis, including declaring MAOIs to be an unfair advantage for Team Old and removing all of them, changing which minor tricylics I do and don’t include in the data, and altering whether Deprenyl, a drug that technically came out in 1970 but received a gritty reboot under the name Emsam in 2006, is counted as older or newer.

So if you want to know what medication will make you happiest, at least according to this analysis your best bet isn’t to ask your doctor, check what’s most popular, or even check any individual online rating database. It’s to look at the approval date on the label and choose the one that came out first.

IV.

What the hell is going on with these data?

I would like to dismiss this as confounded, but I have to admit that any reasonable person would expect the confounders to go the opposite way.

That is: older, less popular drugs are usually brought out only when newer, more popular drugs have failed. MAOIs, the clear winner of this analysis, are very clearly reserved in the guidelines for “treatment-resistant depression”, ie depression you’ve already thrown everything you’ve got at. But these are precisely the depressions that are hardest to treat.

Imagine you are testing the fighting ability of three people via ten boxing matches. You ask Alice to fight a Chihuahua, Bob to fight a Doberman, and Carol to fight Cthulhu. You would expect this test to be biased in favor of Alice and against Carol. But MAOIs and all these other older rarer drugs are practically never brought out except against Cthulhu. Yet they still have the best win-loss record.

Here are the only things I can think of that might be confounding these results.

Perhaps because these drugs are so rare and unpopular, psychiatrists only use them when they have really really good reason. That is, the most popular drug of the year they pretty much cluster-bomb everybody with. But every so often, they see some patient who seems absolutely 100% perfect for clomipramine, a patient who practically screams “clomipramine!” at them, and then they give this patient clomipramine, and she does really well on it.

(but psychiatrists aren’t actually that good at personalizing antidepressant treatments. The only thing even sort of like that is that MAOIs are extra-good for a subtype called atypical depression. But that’s like a third of the depressed population, which doesn’t leave much room for this super-precise-targeting hypothesis.)

Or perhaps once drugs have been on the market longer, patients figure out what they like. Brintellix is so new that the Brintellix patients are the ones whose doctors said “Hey, let’s try you on Brintellix” and they said “Whatever”. MAOIs have been on the market so long that presumably MAOI patients are ones who tried a dozen antidepressants before and stayed on MAOIs because they were the only ones that worked.

(but Prozac has been on the market 25 years now. This should only apply to a couple of very new drugs, not the whole list.)

Or perhaps the older drugs have so many side effects that no one would stay on them unless they’re absolutely perfect, whereas people are happy to stay on the newer drugs even if they’re not doing much because whatever, it’s not like they’re causing any trouble.

(but Seroquel and Abilify, two very new drugs, have awful side effects, yet are down at the bottom along with all the other new drugs)

Or perhaps patients on very rare weird drugs get a special placebo effect, because they feel that their psychiatrist cares enough about them to personalize treatment. Perhaps they identify with the drug – “I am special, I’m one of the only people in the world who’s on nefazodone!” and they become attached to it and want to preach its greatness to the world.

(but drugs that are rare because they are especially new don’t get that benefit. I would expect people to also get excited about being given the latest, flashiest thing. But only drugs that are rare because they are old get the benefit, not drugs that are rare because they are new.)

Or perhaps psychiatrists tend to prescribe the drugs they “imprinted on” in medical school and residency, so older psychiatrists prescribe older drugs and the newest psychiatrists prescribe the newest drugs. But older psychiatrists are probably much more experienced and better at what they do, which could affect patients in other ways – the placebo effect of being with a doctor who radiates competence, or maybe the more experienced psychiatrists are really good at psychotherapy, and that makes the patient better, and they attribute it to the drug.

(but read on…)

V.

Or perhaps we should take this data at face value and assume our antidepressants have been getting worse and worse over the past fifty years.

This is not entirely as outlandish as it sounds. The history of the past fifty years has been a history of moving from drugs with more side effects to drugs with fewer side effects, with what I consider somewhat less than due diligence in making sure the drugs were quite as effective in the applicable population. This is a very complicated and controversial statement which I will be happy to defend in the comments if someone asks.

The big problem is: drugs go off-patent after twenty years. Drug companies want to push new, on-patent medications, and most research is funded by drug companies. So lots and lots of research is aimed at proving that newer medications invented in the past twenty years (which make drug companies money) are better than older medications (which don’t).

I’ll give one example. There is only a single study in the entire literature directly comparing the MAOIs – the very old antidepressants that did best on the patient ratings – to SSRIs, the antidepressants of the modern day4. This study found that phenelzine, a typical MAOI, was no better than Prozac, a typical SSRI. Since Prozac had fewer side effects, that made the choice in favor of Prozac easy.

Did you know you can look up the authors of scientific studies on LinkedIn and sometimes get very relevant information? For example, the lead author of this study has a resume that clearly lists him as working for Eli Lilly at the time the study was conducted (spoiler: Eli Lilly is the company that makes Prozac). The second author’s LinkedIn profile shows he is also an operations manager for Eli Lilly. Googling the fifth author’s name links to a news article about Eli Lilly making a $750,000 donation to his clinic. Also there’s a little blurb at the bottom of the paper saying “Supported by a research grant by Eli Lilly and company”, then thanking several Eli Lilly executives by name for their assistance.

This is the sort of study which I kind of wish had gotten replicated before we decided to throw away an entire generation of antidepressants based on the result.

But who will come to phenelzine’s defense? Not Parke-Davis , the company that made it: their patent expired sometime in the seventies, and then they were bought out by Pfizer5. And not Pfizer – without a patent they can’t make any money off Nardil, and besides, Nardil is competing with their own on-patent SSRI drug Zoloft, so Pfizer has as much incentive as everyone else to push the “SSRIs are best, better than all the rest” line.

Every twenty years, pharmaceutical companies have an incentive to suddenly declare that all their old antidepressants were awful and you should never use them, but whatever new antidepressant they managed to dredge up is super awesome and you should use it all the time. This sort of does seem like the sort of situation that might lead to older medications being better than newer ones. A couple of people have been pushing this line for years – I was introduced to it by Dr. Ken Gillman from Psychotropical Research, whose recommendation of MAOIs and Anafranil as most effective match the patient data very well, and whose essay Why Most New Antidepressants Are Ineffective is worth a read.

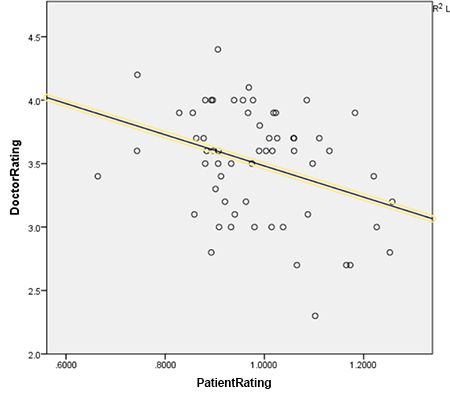
I’m not sure I go as far as he does – even if new antidepressants aren’t worse outright, they might still trade less efficacy for better safety. Even if they handled the tradeoff well, it would look like a net loss on patient rating data. After all, assume Drug A is 10% more effective than Drug B, but also kills 1% of its users per year, while Drug B kills nobody. Here there’s a good case that Drug B is much better and a true advance. But Drug A’s ratings would look better, since dead men tell no tales and don’t get to put their objections into online drug rating sites. Even if victims’ families did give the drug the lowest possible rating, 1% of people giving a very low rating might still not counteract 99% of people giving it a higher rating.

And once again, I’m not sure the tradeoff is handled very well at all.6.

VI.

In order to distinguish between all these hypotheses, I decided to get a lot more data.

I grabbed all the popular antipsychotics, antihypertensives, antidiabetics, and anticonvulsants from the three databases, for a total of 55,498 ratings of 74 different drugs. I ran the same analysis on the whole set.



The three databases still correlate with each other at respectable levels of +0.46, +0.54, and +0.53. All of these correlations are highly significant, p < 0.01.  
  
The negative correlation between patient rating and doctor rating remains and is now a highly significant -0.344, p < 0.01. This is robust even if antidepressants are removed from the analysis, and is notable in both psychiatric and nonpsychiatric drugs.

The correlation between patient rating and year of release is a no-longer-significant -0.191. This is heterogenous; antidepressants and antipsychotics show a strong bias in favor of older medications, and antidiabetics, antihypertensives, and anticonvulsants show a slight nonsignificant bias in favor of newer medications. So it would seem like the older-is-better effect is purely psychiatric.

I conclude that for some reason, there really is a highly significant effect across all classes of drugs that makes doctors love the drugs patients hate, and vice versa.

I also conclude that older psychiatric drugs seem to be liked much better by patients, and that this is not some kind of simple artifact or bias, since if such an artifact or bias existed we would expect it to repeat in other kinds of drugs, which it doesn’t.

VII.

Please feel free to check my results. Here is a spreadsheet (.xls) containing all of the data I used for this analysis. Drugs are marked by class: 1 is antidepressants, 2 is antidiabetics, 3 is antipsychotics, 4 is antihypertensives, and 5 is anticonvulsants. You should be able to navigate the rest of it pretty easily.

One analysis that needs doing is to separate out drug effectiveness versus side effects. The numbers I used were combined satisfaction ratings, but a few databases – most notably WebMD – give you both separately. Looking more closely at those numbers might help confirm or disconfirm some of the theories above.

If anyone with the necessary credentials is interested in doing the hard work to publish this as a scientific paper, drop me an email and we can talk.

Footnotes

1. Technically, MAOI superiority has only been proven for atypical depression, the type of depression where you can still have changing moods but you are unhappy on net. But I’d speculate that right now most patients diagnosed with depression have atypical depression, far more than the studies would indicate, simply because we’re diagnosing less and less severe cases these days, and less severe cases seem more atypical.

2. First-place winner Nardil has only 16% as many reviews as last-place winner Viibryd, even though Nardil has been on the market fifty years and Viibryd for four. Despite its observed superiority, Nardil may very possibly be prescribed less than 1% as often as Viibryd.

3. Pretty much the same thing is true if, instead of looking at the year they came out, you just rank them in order from earliest to latest.

4. On the other hand, what we do have is a lot of studies comparing MAOIs to imipramine, and a lot of other studies comparing modern antidepressants to imipramine. For atypical depression and dysthymia, MAOIs beat imipramine handily, but the modern antidepressants are about equal to imipramine. This strongly implies the MAOIs beat the modern antidepressants in these categories.

5. Interesting Parke-Davis facts: Parke-Davis got rich by being the people to market cocaine back in the old days when people treated it as a pharmaceutical, which must have been kind of like a license to print money. They also worked on hallucinogens with no less a figure than Aleister Crowley, who got a nice tour of their facilities in Detroit.

6. Consider: Seminars In General Psychiatry estimates that MAOIs kill one person per 100,000 patient years. A third of all depressions are atypical. MAOIs are 25 percentage points more likely to treat atypical depression than other antidepressants. So for every 100,000 patients you give a MAOI instead of a normal antidepressant, you kill one and cure 8,250 who wouldn’t otherwise be cured. The QALY database says that a year of moderate depression is worth about 0.6 QALYs. So for every 100,000 patients you give MAOIs, you’re losing about 30 QALYs and gaining about 3,300.

# The Future Is Filters

Related to: The Toxoplasma of Rage

I.

Tumblr Savior is a neat program that blocks Tumblr posts containing specific words or phrases. For example, if you don’t want to hear all of the excellent reasons going around Tumblr why you should kill all men, you just block “kill all men” and they never show up. Add a few extra terms like “white dudes” (nothing good ever came of an article including the phrase “white dudes”), “trans”, “cis”, and “pictures of my vagina”, and you can make Tumblr almost usable.

(My own Tumblr Savior list is an interesting record both of my psyche and of mid-2010s current events. Sometimes I imagine a future cyber-archaeologist stumbling across it and asking “But, but…why would he ban the word ‘puppies’?” Poor, poor innocent future archaeologist.)

I recently learned about Twitter blockbots. These are lists maintained by some trustworthy people, such that subscribing to the blockbot automatically blocks everyone on the list. The original was made by some people in the social justice community to help block people they figured other members of the social justice community wouldn’t want to have to deal with. Although some people seem to be added on by hand, the bot also makes educated guesses about who to block by blacklisting accounts that follow the feeds of too many anti-social-justice leaders.

There are rumors of a similar anti-SJ block list of people who engage on online mobbing and harassment in the name of social justice, but I can’t find it online right now and I think it might have been taken down.

An article I read recently (but which I can’t find right now to link to) proposes a higher-tech solution for Facebook’s harassment problems. They want Facebook to train machine-learning programs to detect posts that most people would consider trollish. So far, so boring. The interesting part comes afterwards – instead of auto-blocking those posts, Facebook would assign them a certain number of Troll Points. Users could then set an option for how their Facebook feed should react to Troll Points – for example, by blocking every post with more than a certain amount. That way, people who were concerned about free speech and who enjoy participating in “heated discussion” would be able to do so, while people who wanted a safer and more pleasant browsing experience could have a very low cutoff for taking action.

But the really interesting part got dismissed after a sentence. What if instead of combining everything into Troll Points, Facebook assigned the points in different domains? Foul Language, Blasphemy, Racial Slurs, Threats, Harassment, Dirty Argument Tactics, et cetera. And then I could set that I don’t care about Foul Language or Blasphemy, but I really don’t want to see any Threats or Racial Slurs.

(obviously the correct anarcho-capitalist solution is to have third-party companies making these algorithms and selling them to individual Facebook users, but in a world where Facebook is trying to become more and more closed to third-party apps, that’s probably not going to happen)

So, take all this filtering technology – Tumblr Savior, Twitter blockbots, and hypothetical Facebook Troll Points, combine them together, project them about ten years into the future with slightly better machine learning, and you have an Internet where nobody has to see, even for an instant, anything they don’t want to. What are the implications?

II.

The most obvious possibility is that everyone will be better off because we can avoid trolls. In this nice black-and-white worldview, there are good people, and there are trolls, and eliminating the trolls is a simple straightforward decision that makes the good people better off. This is how The Daily Beast thinks of it (How Block Bot Could Save The Internet), and as anyone who’s been trolled or harassed online knows, there’s a lot of truth to this view.

The second most obvious possibility is that we will become a civilization of wusses safely protected from ever having to hear an opinion we disagree with, or ever having our prejudices challenged. This is how Reason thinks of it (Block Bots Automate Epistemic Closure On Twitter). Surely there’s some truth here too. How hard would it be to create a filter that blocks all conservative/liberal opinions? Just guess based on whether a text links to foxnews.com or dailykos.com, or add in linguistic cues (“death tax”, “job creators”, etc). Once such a filter existed, how many people do you think would use it proudly, bragging about how they’re no longer “wasting their time listening patiently to bigots” or whatever?

But I don’t think the scenario is quite that apocalyptic. If you’re getting all of your exposure to opinions you disagree with from them being shouted in your face by people you can’t avoid, you probably are not going to lose much by not having that happen. The people who are actually interested in holding discussions can still do that. When I was young and therefore stupid I used to hang out at politics forums specifically for this purpose.

The third possibility is that there would be a remarkable shift of discourse in favor of the powerful and against the powerless.

Terrorism has always been a useful weapon of the powerless. The powerful get laws passed through Congress or whatever, but the powerless don’t have that opportunity. They need to get people to pay attention, and blowing those people up has always been an effective tool in that repertoire. We see this most obviously in places like Palestine and the Basque Country. Likewise, as many people have pointed out, the recent riots in Baltimore can be thought of as a group of powerless people trying to make their anger heard in one of the only ways available to them. It would be politically un-savvy to call this “terrorism”, but as acts of destruction intended to promote a political struggle, they probably fit into the same cluster.

But the next step down from terrorism is annoyism. Terrorism is meant to convince by terrorizing those who ignore your cause; annoyism is meant to convince by annoying people who ignore your cause. Think of a bunch of protesters shouting on a major road, or throwing red paint over people wearing fur, or passive-aggressive Tumblr posts starting “dear white dudes”, or, in probably the purest example of the idea, the Black Brunch protests, where a bunch of black people burst into predominantly white restaurants and shout at patrons about how they’re probably complicit with racism. Even if there’s no implicit threat of force, the point is it’s unpleasant and people can’t ignore it even if they want to.

And so the traditional revolutionary chant goes: “No justice, no peace.” But the thing about filters is that they offer the opportunity for peace regardless of whether or not there is justice. At least they do online, which is where people in the future are going to be spending a lot more of their time.

Imagine you are a rich person who doesn’t want to have to listen to people talking about how rich people need to be socially responsible all the time. It makes you feel guilty, and they are saying mean things like that you don’t deserve all of the money you have, and shouting about social parasites and so on.

So you tell your automated filter to just never let you see any message like that again.

There is an oft-discussed division between politically right or neutral loud angry people (“trolls”) and loud angry people on the political left, (“you are not allowed to dictate the terms on which victims of oppression express their righteous anger”). Machine learning programs will not accept that division, and the latter can be magicked out of visibility just as easily as the former.

Imagine being able to put an entire movement on mute. While I can’t deny the appeal, I’m not sure we – and especially not the social justice community, which is currently laughing at the complaints of people who object to their blockbot – have entirely thought this one through.

III.

The part I find most interesting about all of these possibilities is that they force us to bring previously unconscious social decisions into consciousness.

I think most people, if asked “Is it important to listen to arguments by people who disagree with you?” would answer in the affirmative. I also think most people don’t really do this. Maybe having to set a filter would make people explicitly choose to allow some contrary arguments in. Having done that, people could no longer complain about seeing them – they would feel more of an obligation to read and think about them. And of course, anyone looking for anything more than outrage-bait would choose to preferentially let in high-quality, non-insulting examples of disagreeing views, and so get inspired to think clearly instead of just starting one more rage spiral.

And I think most people, if asked “Is it important to listen to the concerns of the less powerful?” would also be pretty strongly in favor – with the caveat that people can recognize annoyism when it’s being used against them and aren’t especially tolerant of it. The ability to completely block out annoyism, combined with people being forced to explicitly choose to listen to alternative opinions, might make groups that currently favor annoyism change tactics to something more pleasant – though possibly less effective.

I think the result would be several carefully separated groups with their own social and epistemic norms, all of which coexist peacefully and in relative isolation from one another – groups which I would hope then develop their own norms about helping powerless members. This would be an interesting step towards what I describe in my Archipelago article as “a world where everyone is a member of more or less the community they deserve.”

# Growth Mindset 4: Growth Of Office

Previously In Series: No Clarity Around Growth Mindset…Yet // I Will Never Have The Ability To Clearly Explain My Beliefs About Growth Mindset // Growth Mindset 3: A Pox On Growth Your Houses

Last month I criticized a recent paper, Paunesku et al’s Mindset Interventions Are A Scalable Treatment For Academic Underachievement, saying that it spun a generally pessimistic set of findings about growth mindset into a generally optimistic headline.

Earlier today, lead author Dr. Paunesku was kind enough to write a very thorough reply, which I reproduce below:

I.

Hi Scott,

Thanks for your provocative blog post about my work (I’m the first author of the paper you wrote about). I’d like to take a few moments to respond to your critiques, but first I’d like to frame my response and tell you a little bit about my own motivation and that of the team I am a member of (PERTS).

Good criticism is what makes science work. We are critical of our own work, but we are happy to have help. Often critics are not thoughtful or specific. So I very much appreciate the intent of your blog (to be thoughtful and specific).

What is our motivation? We are trying to improve our education system so that all students can thrive. If growth mindset is effective, we want it in every classroom possible. If it is ineffective, we want to know about it so we don’t waste people’s time. If it is effective for some students in some classrooms, we want to know where and for whom so that we can help those students.

What is our history and where are we now? PERTS approached social psychological interventions with a fair amount of skepticism at first. In many ways, they seemed too good to be true. But, we thought, “if this is true, we should do everything we can to spread it”. Our work over the last 5 years has been devoted to trying to see if the results that emerged from initial, small experiments (like Aronson et al., 2002 and Blackwell et al., 2007) would continue to be effective when scaled. The paper you are critiquing is a step in that process — not the end of the process. We are continuing research to see where, for whom, and at what scale social psychological approaches to improving education outcomes can be effective.

How do I intend to respond to your criticisms? In some cases, your facts or interpretations are simply incorrect, and I will try to explain why. I also invite you to contact me for follow up. In others cases, we simply have different opinions about what’s important, and we’ll have to agree to disagree. Regardless, I appreciate your willingness to be bold and specific in your criticism. I think that’s brave, and I think such bravery makes science stronger.

First, what is growth mindset?

This quote is from one of your other blog posts (not your critique of my paper), from your post:

If you’re not familiar with it, growth mindset is the belief that people who believe ability doesn’t matter and only effort determines success are more resilient, skillful, hard-working, perseverant in the face of failure, and better-in-a-bunch-of-other-ways than people who emphasize the importance of ability. Therefore, we can make everyone better off by telling them ability doesn’t matter and only hard work does.

If you think that’s what growth mindset is, I can certainly see why you’d find it irritating — and even destructive. I’d like to assure you that the people doing growth mindset research do not ascribe to the interpretation of growth mindset you described. Nor is that interpretation of growth mindset something we aim to communicate through our interventions. So what is growth mindset?

Growth mindset is not the belief that “ability doesn’t matter and only effort determines success.” Growth mindset is the belief that individuals can improve their abilities — usually through effort and by learning more effective strategies. For example, imagine a third grader struggling to learn long division for the first time. Should he interpret his struggle as a sign that he’s bad at math — as a sign that he should give up on math for good? Or would it be more adaptive if he realized that he could probably get a lot better at math if he sought out help from his peers or teachers? The student who thinks he should give up would probably do pretty badly while the student who thinks that he can improve his abilities — and tries to do so by learning new study strategies and practicing them — would do comparatively better.

That’s the core of growth mindset. It’s nothing crazy like thinking ability doesn’t matter. It’s keeping in mind that you can improve and that — to do so — you need to work hard and seek out and practice new, effective strategies.

As someone who has worked closely with Carol Dweck and with her students and colleagues for seven years now, I can personally attest that I have never heard anyone in that extended group of people express the belief that ability does not matter or that only hard work matters. In fact, a growth mindset wouldn’t make any sense if ability didn’t matter because a growth mindset is all about improving ability.

One of the active goals of the group I co-founded (PERTS) is to try to dispel misinterpretations of growth mindset because they can be harmful. I take it as a failure of our group that someone like you — someone who clearly cares about research and about scientific integrity — could walk away from our work with that interpretation of growth mindset. I hope that PERTS, and other groups promoting growth mindset, can get better and better at refining the way we talk about growth mindset so that people can walk away from our work understanding it more clearly. For that perspective, I hope you can continue to engage with us to improve that message so that people don’t continue to misinterpret it.

Anyway, here are my responses to specific points you made in your blog about my paper:

Was the control group a mindset intervention?

You wrote:

“A quarter of the students took a placebo course that just presented some science about how different parts of the brain do different stuff. This was also classified as a “mindset intervention”, though it seems pretty different.”

What makes you think it was classified as a mindset intervention? We called that the control group, and no one on our team ever thought of that as a mindset intervention.

The Elderly Hispanic Woman Effect

You wrote:

Subgroup analysis can be useful to find more specific patterns in the data, but if it’s done post hoc it can lead to what I previously called the Elderly Hispanic Woman Effect…

First, I just want to note that I love calling this the “elderly Hispanic woman effect.” It really brings out the intrinsic ridiculousness of the subgroup analyses researchers sometimes go through in search of an effect with a p<.05. It is indeed unlikely that “elderly Hispanic women” would be a meaningful subgroup for analyzing the effects of a medicine (although it might be a fun thought exercise to try to think of examples of a medicine whose effects would be likely to be moderated by being an elderly Hispanic woman).

In bringing up the elderly Hispanic woman effect, you’re suggesting that we didn’t have an a priori reason to think that underperforming students would benefit from these mindset interventions and that we just looked through a bunch of moderators until we found one with p<.05. Well that’s not what we did, and I hope I can convince you that our choice of moderator was perfectly reasonable given prior research and theory.

There’s a lot of research (and common sense too) to suggest that mindset — and motivation in general — matters much more when something is hard than when it is easy. Underachieving students presumably find school more difficult, so it makes sense that we’d want to focus on them. I don’t think our choice of subgroup is a controversial or surprising prediction. I think anyone who knows mindset research well would predict stronger effects for students who are struggling. In other words, this is obviously not a case of the elderly Hispanic woman effect because it is totally consistent with prior theory and predictions. What ultimately matters more than any rhetorical argument, however, is whether the effect is robust — whether it replicates.

On that front, I hope you’ll be pleased to learn that we just ran a successful replication of this study (in fall 2014) in which we again found that growth mindset improves achievement specifically among at-risk high school students (currently under review). We’re also planning yet another large scale replication study this fall with a nationally representative sample of schools so that we can be more confident that the interventions are effective in various types of contexts before giving them away for free to any school that wants them.

Is the sense of purpose intervention just a bunch of platitudes?

You wrote:

Still another quarter took a course about “sense of purpose” which talked about how schoolwork was meaningful and would help them accomplish lots of goals and they should be happy to do it.

[Later you say that those “children were told platitudes about how doing well in school will “make their families proud” and “make a positive impact”.]

I wouldn’t say those are platitudes. I think you’re under-appreciating the importance of finding meaning in one’s work. It’s a pretty basic observation about human nature that people are more likely to try hard when it seems like there’s a good reason to try hard. I also think it’s a pretty basic observation about our education system that many students don’t have good reasons for trying hard in school — reasons that resonate with them emotionally and help them find the motivation to do their best in the classroom. In our purpose intervention, we don’t just tell students what to think. We try to scaffold them to think of their own reasons for working hard in school, with a focus on reasons that are more likely to have emotional resonance for students. This type of self-persuasion technique has been used for decades in attitudes research.

We’ve written in more depth about these ideas and explored them through a series of studies. I’d encourage you to read this article if you’re interested.

Our paper title and abstract are misleading

You wrote:

Among ordinary students, the effect on the growth mindset group was completely indistinguishable from zero, and in fact they did nonsignificantly worse than the control group. This was the most basic test they performed, and it should have been the headline of the study. The study should have been titled “Growth Mindset Intervention Totally Fails To Affect GPA In Any Way”.

I think the title you suggest would have been misleading. How?

First, we did find evidence that mindset interventions help underachieving students — and those students are very important from a policy standpoint. As we describe in the paper, those students are more likely to drop out, to end up underemployed, or to end up in prison. So if something can help those students at scale and at a low cost, it’s important for people to know that. That’s why the word “underachievement” is in the title of the paper — because we’re accurately claiming that these interventions can help the important (and large) group of students who are underachieving.

Second, the interventions influenced the way all students think about school in ways that are associated with achievement. Although the higher performing students didn’t show any effects on grades in the semester following the study, their mindsets did change. And, as per the arguments I presented above about the link between mindset and difficulty, it’s quite feasible that those higher-performing students will benefit from this change in mindset down the line. For example, they may choose to take harder classes (e.g., Romero et al., 2014) or they may be more persistent and successful in future classes that are very challenging for them.

A misinterpretation of the y-axis in this graph.

You wrote:

Growth mindset still doesn’t differ from zero [among at-risk students].

This just seems to be a simple misreading of the graph. Either you missed the y-axis of the graph that you reproduced on your blog or you don’t know what a residual standardized score is. Either way, I’ll explain because this is pretty esoteric stuff.

The zero point of the y-axis on that graph is, by definition, the grand mean of the 4 conditions. In other words, the treatment conditions are all hovering around zero because zero is the average, and the average is made up mostly of treatment group students. If we had only had 2 conditions (each with 50% of the students), the y-axis “zero” would have been exactly halfway in between them. So the lack of difference from zero does not mean that the treatment was not different from control. The relevant comparison is between the error bars in the control condition and in the treatment conditions.

You might ask, “why are you showing such a graph?” We’re doing so to focus on the treatment contrast at the heart of our paper — the contrast between the control and treatment groups. The residual standardized graph makes it easy to see the size of that treatment contrast.

We’re combining intervention conditions

You wrote:

Did you catch that phrase “intervention conditions”? The authors of the study write: “Because our primary research question concerned the efficacy of academic mindset interventions in general when delivered via online modules, we then collapsed the intervention conditions into a single intervention dummy code (0 = control, 1 = intervention).

[This line of argument goes on for a long time to suggest that we’re unethical and that there’s actually no evidence for the effects of growth mindset on achievement.]

We collapsed the intervention conditions together for this analysis because we were interested in the overall effect of these interventions on achievement. We wanted to see if it is possible to use scalable, social-psychological approaches to improve the achievement of underperforming students. I’m not sure why you think that’s not a valid hypothesis to test, but we certainly think it is. Maybe this is just a matter of opinion about what’s a meaningful hypothesis to test, but I assure you that this hypothesis (contrast all treatments to control) is consistent with the goal of our group to develop treatments that make an impact on student achievement. As I described before, we have a whole center devoted to trying to improve academic achievement with these types of techniques (see perts.net); so it’s pretty natural that we’d want to see whether our social-psychological interventions improve outcomes for the students who need them most (at-risk students).

You’re correct that the growth mindset intervention did not have a statistically significant impact on course passing rates by itself (at a p<.05 level). However, the effect was in the expected direction with p=0.13 (or a 1-tailed p=.07 — I hope you’ll grant that a 1-tailed test is appropriate here given that we obviously predicted the treatment would improve rather than reduce performance). So the lack of a p<.05 should not be interpreted — as you seem to interpret it — as some sort of positive evidence that growth mindset “actually didn’t work.” Anyway, I would say it warrants further research to replicate this effect (work we are currently engaging in).

To summarize, we did not find direct evidence that the growth mindset intervention increased course passing rates on its own at a p<.05 level. We did find that growth mindset increased course passing rates at a trend level — and found a significant effect on GPA. More importantly for me (though perhaps less relevant to your interest specifically in growth mindset), we did provide evidence that social-psychological interventions, like growth mindset and sense of purpose, can improve academic outcomes for at-risk students.

We’re excited to be replicating this work now and giving it away in the hopes of improving outcomes for students around the world.

Summary

I hope I addressed your concerns about this paper, and I welcome further discussion with you. I’d really appreciate it if you’d revise your blog post in whatever way you think is appropriate in light of my response. I’d hate for people to get the wrong impression of our work, and you don’t strike me as someone who would want to mislead people about scientific findings either.

Finally, you’re welcome to post my response. I may post it to my own web page because I’m sure many other people have similar questions about my work. Just let me know how you’d like to proceed with this dialog.

Thanks for reading,

Dave

II.

First of all, the obvious: this is extremely kind and extremely well-argued and a lot of it is correct and makes me feel awful for being so snarky on my last post.

Things in particular which I want to endorse as absolutely right about the critique:

I wrote “A quarter of the students took a placebo course that just presented some science about how different parts of the brain do different stuff. This was also classified as a “mindset intervention”, though it seems pretty different.” Dr. Paunesku says this is wrong. He’s right. It was an editing error on my part. I meant to add the last sentence to the part on the “sense of purpose” intervention, which was classified as a mindset intervention and which I do think seems pretty different. The placebo intervention was never classified as a mindset intervention and I completely screwed up by inserting that piece of text there rather than two sentences down where I meant it to be. It has since been corrected and I apologize for the error.

If another successful replication found that growth mindset continues to only help the lowest-performing students, I withdraw the complaint that this is sketchy subgroup mining, though I think that in general worrying about this is the correct thing to do.

I did misunderstand the residual standardized graph. I suggested that the control group must have severely declined, and got confused about why. In fact, the graph was not about difference between pre-study scores and post-study scores, but difference between group scores and the average score for all four groups. So when the control group is strongly negative, that means it was much worse than the average of all groups. When growth mindset is not-different-from-zero, it means growth mindset was not different from the average of all four groups, which consists of three treatment groups and one control group. So my interpretation – that growth mindset failed to change children’s grades – is not supported by the data.

(In my defense, I can only plead that in the two hundred fifty comments I received, many by professional psychologists and statisticians, only one person picked up on this point (admittedly, after being primed by my own misinterpretation). And the sort of data I expected to be seeing – difference between students’ pre-intervention and post-intervention scores – does not seem to be available. Nevertheless, this was a huge and unforgiveable screw-up, and I apologize.)

III.

But there are also a few places where I will stick to my guns.

I don’t think my interpretation of growth mindset was that far off the mark. I explain this a little further in this post on differing possible definitions of growth mindset, and I will continue to cite this strongly worded paper by Dweck as defense of my views. It’s not just an obvious and innocuous belief about about always believing you should be able to improve, it’s a belief about very counterintuitive effects of believing that success depends on ability versus effort. It is possible that all sophisticated researchers in the field have a very sophisticated and unobjectionable definition of growth mindset, but that’s not the way it’s presented to the public, even in articles by those same researchers.

Although I’m sure that to researchers in the field statements like “Doing well at school will help me achieve my goal” don’t sound like platitudes, it seems important to me in the context of discussions about growth mindset. Some people have billed growth mindset as a very exciting window into what makes learning tick, and how we should divide everyone into groups based on their mindset, and how it’s the Secret To Success, and so on. Learning that a drop-dead simple intervention – telling students to care about school more – actually does as well or better than growth mindset seems to me like a damning result. I realize it would be kind of insulting to call sense-of-purpose an “active placebo” in the medical sense, but that’s kind of how I can’t help thinking of it.

I’m certainly not suggesting the authors of the papers are unethical for combining growth mindset intervention with sense of purpose intervention. But I think the technique is dangerous, and this is an example. They got a result that was significant at p = 0.13. Dr. Paunesku suggests in his email to me that this should be one-tailed (which makes it p = 0.07) and that this obviously trends towards significance. This is a reasonable argument. But this wasn’t the reasonable argument made in the paper. Instead, they make it look like it achieved classical p < 0.05 significance, or at least make it very hard to notice that it didn’t.

Even if in this case it was – I can’t even say white lie, maybe a white spin – I find the technique very worrying. Suppose I want to prove homeopathy cures cancer. I make a trial with one placebo condition and two intervention conditions – chemotherapy and homeopathy. I find that the chemotherapy condition very significantly outperforms placebo, but the homeopathy condition doesn’t. So I combine the two interventions into a single bin and say “Therapeutic interventions such as chemotherapy or homeopathy significantly outperform placebo.” Then someone else cites it as “As per a study, homeopathy outperforms placebo.” This would obviously be bad.

I am just not convinced that growth mindset and sense of purpose are similar enough that you can group them together effectively. This is what I was trying to get at in my bungled sentence about how they’re both “mindset” interventions but seem pretty different. Yes, they’re both things you tell children in forty-five minute sessions that seem related to how they think about school achievement. But that’s a really broad category.

But doesn’t it mean something that growth-mindset was obviously trending toward significance?

First of all, I would have had no problem with saying “trending toward significance” and letting readers draw their own conclusions.

Second of all, I’m not totally sure I buy the justification for a one-tailed test here; after all, it seems like we should use a one-tailed test for homeopathy as well, since as astounding as it would be if homeopathy helped, it would be even more astounding if homeopathy somehow made cancer worse. Further, educational interventions often have the opposite of their desired effect – see eg this campaign to increase tolerance of the disabled which made students like disabled people less than a control intervention. In fact, there’s no need to look further than this very study, which found (counterintuitively) that among students already exposed to sense-of-purpose interventions, adding on an extra growth-mindset intervention seemed to make them do (nonsignificantly) worse. I am not a statistician, but my understanding is you ought to have a super good reason to use a one-tailed test, beyond just “Intuitively my hypothesis is way more likely than the exact opposite of my hypothesis”.

Third of all, if we accept p < 0.13 as “trending towards significance”, we have basically tripled the range of acceptable study results, even though everyone agrees our current range of acceptable study results is already way too big and some high percent of all medical studies are wrong and only 39% of psych studies replicate and so on.

(I agree that all of this could be solved by something better than p-values, but p-values are what we’ve got)

I realize I’m being a jerk by insisting on the arbitrary 0.05 criterion, but in my defense, the time when only 39% of studies using a criterion replicate is a bad time to loosen that criterion.

IV.

Here’s what I still believe and what I’ve changed my mind on based on Dr. Paunesku’s response.

1. I totally bungled my sentence on the placebo group being a mindset intervention by mistake. I ashamedly apologize, and have corrected the original post.

2. I totally bungled reading the residual standard score graph. I ashamedly apologize, and have corrected the original post, and put a link in bold text to this post on the top.

3. I don’t know whether the thing I thought the graph showed (no significant preintervention vs. postintervention GPA improvement for growth mindset, or no difference in change from controls) is true. It may be hidden in the supplement somewhere, which I will check later. Possible apology pending further investigation.

4. Growth mindset still had no effect (in fact nonsignificantly negative) for students at large (as opposed to underachievers). I regret nothing.

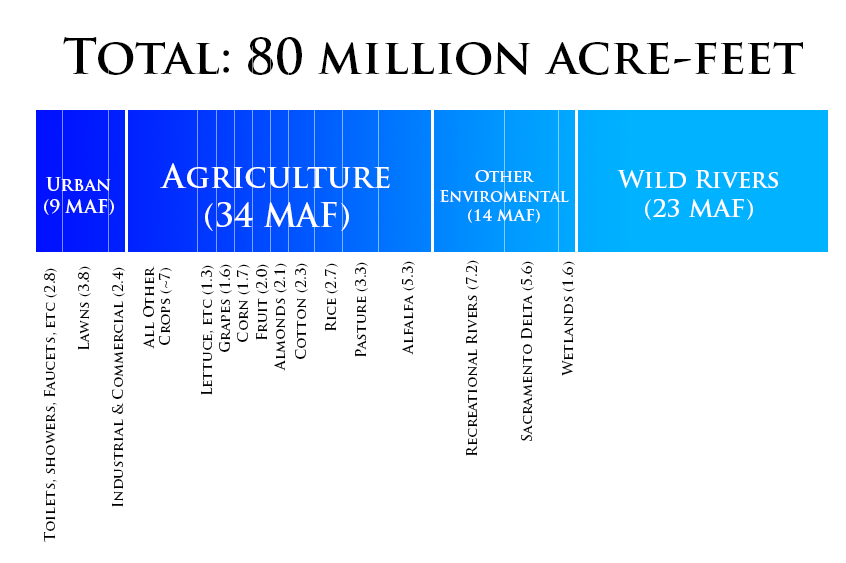
5. Growth mindset still failed to reach traditional significance criteria for changing pass rates. I regret nothing.

# California, Water You Doing?

[Epistemic status: Low confidence. I have found numbers and stared at them until they made sense to me, but I have no education in this area. Tell me if I’m wrong.]

I.

There has recently been a lot of dumb fighting over who uses how much water in California, so I thought I would see if it made more sense as an infographic sort of thing:



Sources include Understanding Water Use In California, Inputs To Farm Production, California Water Usage In Crops, Urban Water Use Efficiency, Water Use In California, and Water: Who Uses How Much. There are some contradictions, probably caused by using sources from different years, and although I’m pretty confident this is right on an order of magnitude scale I’m not sure about a percentage point here or there. But that having been said:

On a state-sized level, people measure water in acre-feet, where an acre-foot is the amount of water needed to cover an area of one acre to a depth of one foot. California receives a total of 80 million acre-feet of water per year. Of those, 23 million are stuck in wild rivers (the hydrological phenomenon, not the theme park). These aren’t dammed and don’t have aqueducts to them so they can’t be used for other things. There has been a lot of misdirection over this recently, since having pristine wild rivers that fish swim in seems like an environmental cause, and so you can say that “environmentalists have locked up 23 million acre-feet of California water”. This is not a complete lie; if not for environmentalism, maybe some of these rivers would have been dammed up and added to the water system. But in practice you can’t dam every single river and most of these are way off in the middle of nowhere far away from the water-needing population. People’s ulterior motives shape whether or not they add these to the pot; I’ve put them in a different color blue to mark this.

Aside from that, another 14 million acre-feet are potentially usable, but deliberately diverted to environmental or recreational causes. These include 7.2 million for “recreational rivers”, apparently ones that people like to boat down, 1.6 million to preserve wetlands, and 5.6 million to preserve the Sacramento River Delta. According to environmentalists, this Sacramento River Delta water is non-negotiable, because if we stopped sending fresh water there the entire Sacramento River delta would turn salty and it would lead to some kind of catastrophe that would threaten our ability to get fresh water into the system at all.

34 million acre-feet of water are diverted to agriculture. The most water-expensive crop is alfalfa, which requires 5.3 million acre-feet a year. If you’re asking “Who the heck eats 5.3 million acre-feet of alfalfa?” the answer is “cows”. A bunch of other crops use about 2 million acre-feet each.

All urban water consumption totals 9 million acre-feet. Of those, 2.4 million are for commercial and industrial institutions, 3.8 million are for lawns, and 2.8 million are personal water use by average citizens in their houses. In case you’re wondering about this latter group, by my calculations all water faucets use 0.5 million, all toilets use 0.9 million, all showers use 0.5 million, leaks lose 0.3 million, and the remaining 0.6 million covers everything else – washing machines, dishwashers, et cetera.

Since numbers like these are hard to think about, it might be interesting to put them in a more intuitive form. The median California family earns $70,000 a year – let’s take a family just a little better-off than that who are making $80,000 so we can map it on nicely to California’s yearly water income of 80 million acre-feet.

The unusable 23 million acre-feet which go into wild rivers and never make it into the pot correspond to the unusable taxes the California family will have to pay. So our family is left with $57,000 post-tax income.

In this analogy, California is spending $14,000 on environment and recreation, $34,000 on agriculture, and $9,000 on all urban areas. All household uses – toilets, showers, faucets, etc – only add up to about $2,800 of their budget.

There is currently a water shortfall of about 6 million acre-feet per year, which is being sustained by exploiting non-renewable groundwater and other sources. This is the equivalent of our slightly-richer-than-average family having to borrow $6,000 from the bank each year to get by.

II.

Armed with this information, let’s see what we can make of some recent big news stories.

Apparently we are supposed to be worried about fracking depleting water in California. ThinkProgress reports that Despite Historic Drought, California Used 70 Million Gallons Of Water For Fracking Last Year. Similar concerns are raised by RT, Huffington Post, and even The New York Times. But 70 million gallons equals 214 acre-feet. Remember, alfalfa production uses 5.3 million acre feet. In our family-of-four analogy above, all the fracking in California costs them about a quarter. Worrying over fracking is like seeing an upper middle class family who are $6,000 in debt, and freaking out because one of their kids bought a gumball from a machine.

Apparently we are also supposed to be worried about Nestle bottling water in California. ABC News writes an article called Nestle Needs To Stop Bottling Water In Drought-Stricken California, Advocacy Group Says, about a group called the “Courage Campaign” who have gotten 135,000 signatures on a petition saying that Nestle needs to stop “bottling the scarce resource straight from the heart of California’s drought and selling it for profit.” Salon goes even further – their article is called Nestle’s Despicable Water Crisis Profiteering: How It’s Making A Killing While California Is Dying Of Thirst, and as always with this sort of thing Jezebel also has to get in on the action. But Nestle’s plant uses only 150 acre-feet, about one forty-thousandth the amount used to grow alfalfa, and the equivalent of about a dime to our family of four.

The Wall Street Journal says that farms are a scapegoat for the water crisis, because in fact the real culprits are environmentalists. They say that “A common claim is that agriculture consumes about 80% of ‘developed’ water supply, yet this excludes the half swiped off the top for environmental purposes.” But environmentalism only swipes half if you count among that half all of the wild rivers in the state – that is, every drop of water not collected, put in an aqueduct, and used to irrigate something is a “concession” to environmentalists. A more realistic figure for environmental causes is the 14 million acre-feet marked “Other Environmental” on the map above, and even that includes concessions to recreational boaters and to whatever catastrophe is supposed to happen if we can’t keep the Sacramento Delta working properly. It’s hard to calculate exactly how much of California’s water goes to environmental causes, but half is definitely an exaggeration.

Wired is concerned that the federal government is ordering California to spend 12,000 acre-feet of water to save six fish (h/t Alyssa Vance). Apparently these are endangered fish in some river who need to get out to the Pacific to breed, and the best way to help them do that is to fill up the river with 12,000 acre feet of water. That’s about $12 on our family’s budget, which works out to $2 per fish. I was going to say that I could totally see a family spending $2 on a fish, especially if it was one of those cool glow-in-the-dark fish I used to have when I was a kid, but then I remembered this was a metaphor and the family is actually the entire state budget of California but the six fish are still literally just six fish. Okay, yes, that seems a little much.

III.

Finally, Marginal Revolution and even some among the mysterious and endangered population of non-blog-having economists are talking about how really the system of price controls and subsidies in the water market is ridiculous and if we had a free market on water all of our problems would be solved. It looks to me like that’s probably right.

Consider: When I used to live in California, even before this recent drought I was being told to take fewer showers, to install low-flush toilets that were inconvenient and didn’t really work all that well, to limit my use of the washing machine and dishwasher, et cetera. It was actually pretty inconvenient. I assume all forty million residents of California were getting the same message, and that a lot of them would have liked to be able to pay for the right to take nice long relaxing showers.

But if all the savings from water rationing amounted to 20% of our residential water use, then that equals about 0.5 MAF, which is about 10% of the water used to irrigate alfalfa. The California alfalfa industry makes a total of $860 million worth of alfalfa hay per year. So if you calculate it out, a California resident who wants to spend her fair share of money to solve the water crisis without worrying about cutting back could do it by paying the alfalfa industry $2 to not grow $2 worth of alfalfa, thus saving as much water as if she very carefully rationed her own use.

If you were to offer California residents the opportunity to not have to go through the whole gigantic water-rationing rigamarole for $2 a head, I think even the poorest people in the state would be pretty excited about that. My mother just bought and installed a new water-saving toilet – which took quite a bit of her time and money – and furthermore, the government is going to give her a $125 rebate for doing so. Cutting water on the individual level is hard and expensive. But if instead of trying to save water ourselves, we just paid the alfalfa industry not to grow alfalfa, all the citizens of California could do their share for $2. If they also wanted to have a huge lush water-guzzling lawn, their payment to the alfalfa industry would skyrocket all the way to $5 per year.

In fact, though I am not at all sure here and I’ll want a real economist to double-check this, it seems to me if we wanted to buy out all alfalfa growers by paying them their usual yearly income to just sit around and not grow any alfalfa, that would cost $860 million per year and free up 5.3 million acre-feet, ie pretty much our entire shortfall of 6 million acre-feet, thus solving the drought. Sure, 860 million dollars sounds like a lot of money, but note that right now California newspapers have headlines like Billions In Water Spending Not Enough, Officials Say. Well, maybe that’s because you’re spending it on giving people $125 rebates for water-saving toilets, instead of buying out the alfalfa industry. I realize that paying people subsidies to misuse water to grow unprofitable crops, and then offering them countersubsidies to not take your first set of subsidies, is to say the least a very creative way to spend government money – but the point is it is better than what we’re doing now.

# Bicameral Reasoning

[Epistemic status: Probably not the first person to think about this, possibly just reinventing scope insensitivity. Title with apologies to Julian Jaynes]

Non-American readers may not be familiar with the history of the US House and Senate.

During the Constitutional Convention, a fight broke out between the smaller states and the bigger states. The smaller states, like Delaware, wanted each state to elect a fixed number of representatives to the legislature, so that Delaware would have just as much of a say as, for example, New York. The bigger states wanted legislative representation to be proportional to population, so that if New York had ten times as many people as Delaware, they would get ten times as many representatives.

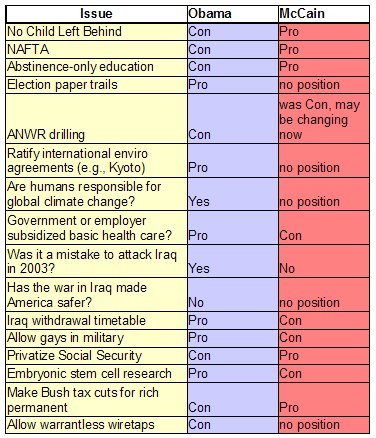
Eventually everyone just agreed to compromise by splitting the legislature into the House of Representatives and the Senate. The House worked the way New York wanted things, the Senate worked the way Delaware wanted things, and they would have to agree to get anything done.

This system has continued down to the present. Today, Delaware has only one Representative, far less than New York’s twenty-seven. But both states have an equal number of Senators, even though New York has a population of twenty million and Delaware is uninhabited except by corporations looking for tax loopholes.

To me, the House system seems much fairer. If New York has ten times the population of Delaware, but both have the same number of representatives, then Delaware citizens have ten times as much political power just because they live on one side of an arbitrary line. And New York might be tempted to split up into ten smaller states, and thus increase its political power tenfold. Heck, why don’t we just declare some random farm a state and give five people and a cow the same political power as all of California?

But despite my professed distaste for the Senate’s representational system, I find myself using something similar in parts of my own thought processes where I least expect.

Every election, I see charts like this:



And I tend to think something like “Well, I agree with this guy about the Iraq war and global warming, but I agree with that guy about election paper trails and gays in the military, so it’s kind of a toss-up.”

And this way of thinking is awful.

The Iraq War probably killed somewhere between 100,000 and 1,000,000 people. If you think that it was unnecessary, and that it was possible to know beforehand how poorly it would turn out, then killing a few hundred thousand people is a really big deal. I like having paper trails in elections as much as the next person, but if one guy isn’t going to keep a very good record of election results, and the other guy is going to kill a million people, that’s not a toss-up.

Likewise with global warming versus gays in the military. It would be nice if homosexual people have the same right to be killed by roadside explosive devices that the rest of us enjoy, but not frying the planet is pretty important too.

(if you don’t believe in global warming, fine, having a government that agrees with you and doesn’t waste 5% of the world GDP fighting it is still more important than anything else on this list)

Saying “some boxes are more important than others” doesn’t really cut it; it sounds like they might be twice, maybe three times more important, whereas in fact they might literally be a million times more important. It doesn’t convey the right sense of “Why are you even looking at that other box?”

I worry that, by portraying issues in this nice little set of boxes, this graphic is priming reasoning similar to the US Senate, where each box gets the same level of representation in my decision-making process, regardless of whether it’s a Delaware-sized box that affects a handful of people, or a New York sized box with millions of lives hanging in the balance.

I was thinking about this again back in March when I had a brief crisis caused by worrying that the moral value of the world’s chickens vastly exceeded the moral value of the world’s humans. I ended up being trivially wrong – there are only about twenty billion chickens, as opposed to the hundreds of billions I originally thought. But I was contingently wrong – in other words, I got lucky. Honestly, I didn’t know whether there were twenty billion chickens or twenty trillion.

And honestly, 99% of me doesn’t care. I do want to improve chickens, and I do think that their suffering matters. But thanks to the miracle of scope insensitivity, I don’t particularly care more about twenty trillion chickens than twenty billion chickens.

Once again, chickens seem to get two seats to my moral Senate, no matter how many of them there are. Other groups that get two seats include “starving African children”, “homeless people”, “my patients in hospital”, “my immediate family”, and “my close friends”. Obviously some of these groups contain thousands of times more people than others. They still get two seats. And so I am neither willing to reduce chickens’ values to zero value units per chicken, nor accept that if there are enough chickens they will end up able to outvote everyone else.

(I’m not sure whether “chickens” and “cows” are two separate states, or if there’s just one state of “Animals”. It probably depends on my mood. Which is worrying.)

And most recently I thought about this because of the post on California water I wrote last week. It seems very wise to say we all have to make sacrifices, and to concentrate about equally on natural categories of water use like showers, and toilets, and farms, and lawns – without noticing that one of those is ten times bigger than the other three combined. It seems like most people who think about the water crisis are using a Senate model, where each category is treated as an equally important area to optimize. In a House model, you wouldn’t be thinking about showers any more than a 2008 voter should be thinking of election paper trails.

I’m tempted to say “The House is just plain right and the Senate is just plain wrong”, but I’ve got to admit that would clash with my own very strong inclinations on things like the chicken problem. The Senate view seems to sort of fit with a class of solutions to the dust specks problem where after the somethingth dust speck or so you just stop caring about more of them, with the sort of environmentalist perspective where biodiversity itself is valuable, and with the Leibnizian answer to Job.

But I’m pretty sure those only kick in at the extremes. Take it too far, and you’re just saying the life of a Delawarean is worth twenty-something New Yorkers.

# Beware Summary Statistics

Last night I asked Tumblr two questions that had been bothering me for a while and got some pretty good answers.

I.

First, consider the following paragraph from JRank:

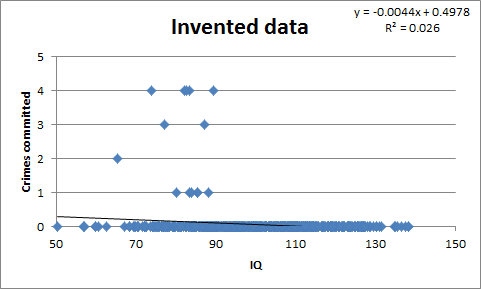
Terrie Moffitt and colleagues studied 4,552 Danish men born at the end of World War II. They examined intelligence test scores collected by the Danish army (for screening potential draftees) and criminal records drawn from the Danish National Police Register. The men who committed two or more criminal offenses by age twenty had IQ scores on average a full standard deviation below nonoffenders, and IQ and criminal offenses were significantly and negatively correlated at r = -.19.

Repeat offenders are a 15 IQ points – an entire standard deviation – below the rest of the population. This matches common sense, which suggests that serial criminals are not the brightest members of society. It sounds from this like IQ is a very important predictor of crime.

But r = – 0.19 suggests that only about 3.6% of variance in crime is predicted by IQ. 3.6% is nothing. It sounds from this like IQ barely matters at all in predicting crime.

This isn’t a matter of conflicting studies: these are two ways of describing the same data. What gives?

The best answer I got was from pappubahry2, who posted the following made-up graph:

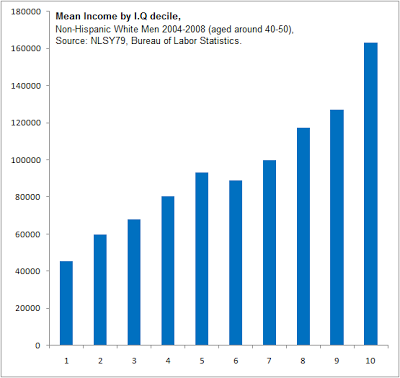


Here all crime is committed by low IQ individuals, but the correlation between IQ and crime is still very low, r = 0.16. The reason is simple: very few people, including very few low-IQ people, commit crimes. r is kind of a mishmash of p(low IQ|criminal) and p(criminal|low IQ), and the latter may be very low even when all criminals are from the lower end of the spectrum.

The advice some people on Tumblr gave was to beware summary statistics. “IQ only predicts 3.6% of variance in crime” makes it sound like IQ is nearly irrelevant to criminality, but in fact it’s perfectly consistant with IQ being a very strong predictive factor.

II.

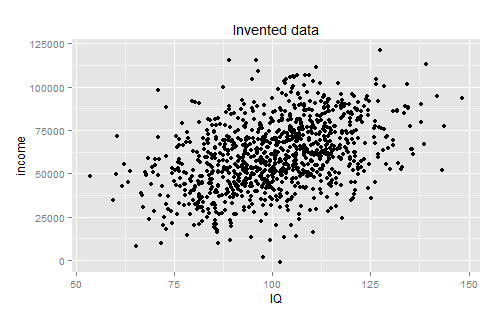
So I pressed my luck with the following question:

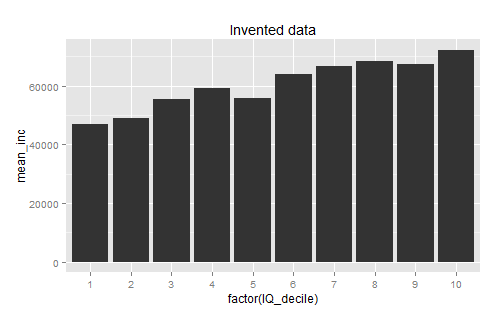


I’m not sure why everyone’s income on this graph is so much higher than average US per capita of $30,000ish, or even average white male income of $31,000ish. I think it might be the ‘age 40 to 50’ specifier.

This graph suggests IQ is an important determinant of income. But most studies say the correlation between IQ and income is at most 0.4 or so, or 16% of the variance, suggesting it’s a very minor determinant of income. Most people are earning an income, so the too-few-criminals explanation from above doesn’t apply. Again, what gives?

The best answer I got for this one was from su3su2u1, who pointed out that there was probably very high variance within the individual deciles. Pappubahry made some more graphs to demonstrate:





I understand this one intellectually, but I still haven’t gotten my head around it. Regardless of the amount of variance, going from a category where I can expect to make on average $40,000 to a category where I could expect to make on average $160,000 seems like a pretty big deal, and describing it as “only predicting 16% of the variation” seems patently unfair.

I guess the moral is the same as the moral in the first situation: beware summary statistics. Based on the way you explain things, you can use different summary statistics to make things look very important or not important at all. And as a bunch of people recommended to me: when in doubt, demand to see the scatter plot.

# AI Researchers On AI Risk

I first became interested in AI risk back around 2007. At the time, most people’s response to the topic was “Haha, come back when anyone believes this besides random Internet crackpots.”

Over the next few years, a series of extremely bright and influential figures including Bill Gates, Stephen Hawking, and Elon Musk publically announced they were concerned about AI risk, along with hundreds of other intellectuals, from Oxford philosophers to MIT cosmologists to Silicon Valley tech investors. So we came back.

Then the response changed to “Sure, a couple of random academics and businesspeople might believe this stuff, but never real experts in the field who know what’s going on.”

Thus pieces like Popular Science’s Bill Gates Fears AI, But AI Researchers Know Better:

When you talk to A.I. researchers—again, genuine A.I. researchers, people who grapple with making systems that work at all, much less work too well—they are not worried about superintelligence sneaking up on them, now or in the future. Contrary to the spooky stories that Musk seems intent on telling, A.I. researchers aren’t frantically installed firewalled summoning chambers and self-destruct countdowns.

And Fusion.net’s The Case Against Killer Robots From A Guy Actually Building AI:

Andrew Ng builds artificial intelligence systems for a living. He taught AI at Stanford, built AI at Google, and then moved to the Chinese search engine giant, Baidu, to continue his work at the forefront of applying artificial intelligence to real-world problems. So when he hears people like Elon Musk or Stephen Hawking—people who are not intimately familiar with today’s technologies—talking about the wild potential for artificial intelligence to, say, wipe out the human race, you can practically hear him facepalming.

And now Ramez Naam of Marginal Revolution is trying the same thing with What Do AI Researchers Think Of The Risk Of AI?:

Elon Musk, Stephen Hawking, and Bill Gates have recently expressed concern that development of AI could lead to a ‘killer AI’ scenario, and potentially to the extinction of humanity. None of them are AI researchers or have worked substantially with AI that I know of. What do actual AI researchers think of the risks of AI?

It quotes the same couple of cherry-picked AI researchers as all the other stories – Andrew Ng, Yann LeCun, etc – then stops without mentioning whether there are alternate opinions.

There are. AI researchers, including some of the leaders in the field, have been instrumental in raising issues about AI risk and superintelligence from the very beginning. I want to start by listing some of these people, as kind of a counter-list to Naam’s, then go into why I don’t think this is a “controversy” in the classical sense that dueling lists of luminaries might lead you to expect.

The criteria for my list: I’m only mentioning the most prestigious researchers, either full professors at good schools with lots of highly-cited papers, or else very-well respected scientists in industry working at big companies with good track records. They have to be involved in AI and machine learning. They have to have multiple strong statements supporting some kind of view about a near-term singularity and/or extreme risk from superintelligent AI. Some will have written papers or books about it; others will have just gone on the record saying they think it’s important and worthy of further study.

If anyone disagrees with the inclusion of a figure here, or knows someone important I forgot, let me know and I’ll make the appropriate changes:

\* \* \* \* \* \* \* \* \* \*

Stuart Russell (wiki) is Professor of Computer Science at Berkeley, winner of the IJCAI Computers And Thought Award, Fellow of the Association for Computing Machinery, Fellow of the American Academy for the Advancement of Science, Director of the Center for Intelligent Systems, Blaise Pascal Chair in Paris, etc, etc. He is the co-author of Artificial Intelligence: A Modern Approach, the classic textbook in the field used by 1200 universities around the world. On his website, he writes:

The field [of AI] has operated for over 50 years on one simple assumption: the more intelligent, the better. To this must be conjoined an overriding concern for the benefit of humanity. The argument is very simple:

1. AI is likely to succeed.  
2. Unconstrained success brings huge risks and huge benefits.  
3. What can we do now to improve the chances of reaping the benefits and avoiding the risks?

Some organizations are already considering these questions, including the Future of Humanity Institute at Oxford, the Centre for the Study of Existential Risk at Cambridge, the Machine Intelligence Research Institute in Berkeley, and the Future of Life Institute at Harvard/MIT. I serve on the Advisory Boards of CSER and FLI.

Just as nuclear fusion researchers consider the problem of containment of fusion reactions as one of the primary problems of their field, it seems inevitable that issues of control and safety will become central to AI as the field matures. The research questions are beginning to be formulated and range from highly technical (foundational issues of rationality and utility, provable properties of agents, etc.) to broadly philosophical.

He makes a similar point on edge.org, writing:

As Steve Omohundro, Nick Bostrom, and others have explained, the combination of value misalignment with increasingly capable decision-making systems can lead to problems—perhaps even species-ending problems if the machines are more capable than humans. Some have argued that there is no conceivable risk to humanity for centuries to come, perhaps forgetting that the interval of time between Rutherford’s confident assertion that atomic energy would never be feasibly extracted and Szilárd’s invention of the neutron-induced nuclear chain reaction was less than twenty-four hours.

He has also tried to serve as an ambassador about these issues to other academics in the field, writing:

What I’m finding is that senior people in the field who have never publicly evinced any concern before are privately thinking that we do need to take this issue very seriously, and the sooner we take it seriously the better.

David McAllester (wiki) is professor and Chief Academic Officer at the U Chicago-affilitated Toyota Technological Institute, and formerly served on the faculty of MIT and Cornell. He is a fellow of the American Association of Artificial Intelligence, has authored over a hundred publications, has done research in machine learning, programming language theory, automated reasoning, AI planning, and computational linguistics, and was a major influence on the algorithms for famous chess computer Deep Blue. According to an article in the Pittsburgh Tribune Review:

Chicago professor David McAllester believes it is inevitable that fully automated intelligent machines will be able to design and build smarter, better versions of themselves, an event known as the Singularity. The Singularity would enable machines to become infinitely intelligent, and would pose an ‘incredibly dangerous scenario’, he says.

On his personal blog Machine Thoughts, he writes:

Most computer science academics dismiss any talk of real success in artificial intelligence. I think that a more rational position is that no one can really predict when human level AI will be achieved. John McCarthy once told me that when people ask him when human level AI will be achieved he says between five and five hundred years from now. McCarthy was a smart man. Given the uncertainties surrounding AI, it seems prudent to consider the issue of friendly AI…

The early stages of artificial general intelligence (AGI) will be safe. However, the early stages of AGI will provide an excellent test bed for the servant mission or other approaches to friendly AI. An experimental approach has also been promoted by Ben Goertzel in a nice blog post on friendly AI. If there is a coming era of safe (not too intelligent) AGI then we will have time to think further about later more dangerous eras.

He attended the AAAI Panel On Long-Term AI Futures, where he chaired the panel on Long-Term Control and was described as saying:

McAllester chatted with me about the upcoming ‘Singularity’, the event where computers out think humans. He wouldn’t commit to a date for the singularity but said it could happen in the next couple of decades and will definitely happen eventually. Here are some of McAllester’s views on the Singularity. There will be two milestones: Operational Sentience, when we can easily converse with computers, and the AI Chain Reaction, when a computer can bootstrap itself to a better self and repeat. We’ll notice the first milestone in automated help systems that will genuinely be helpful. Later on computers will actually be fun to talk to. The point where computer can do anything humans can do will require the second milestone.

Hans Moravec (wiki) is a former professor at the Robotics Institute of Carnegie Mellon University, namesake of Moravec’s Paradox, and founder of the SeeGrid Corporation for industrial robotic visual systems. His Sensor Fusion in Certainty Grids for Mobile Robots has been cited over a thousand times, and he was invited to write the Encyclopedia Britannica article on robotics back when encyclopedia articles were written by the world expert in a field rather than by hundreds of anonymous Internet commenters.

He is also the author of Robot: Mere Machine to Transcendent Mind, which Amazon describes as:

In this compelling book, Hans Moravec predicts machines will attain human levels of intelligence by the year 2040, and that by 2050, they will surpass us. But even though Moravec predicts the end of the domination by human beings, his is not a bleak vision. Far from railing against a future in which machines rule the world, Moravec embraces it, taking the startling view that intelligent robots will actually be our evolutionary heirs.” Moravec goes further and states that by the end of this process “the immensities of cyberspace will be teeming with unhuman superminds, engaged in affairs that are to human concerns as ours are to those of bacteria”.

Shane Legg is co-founder of DeepMind Technologies (wiki), an AI startup that was bought for Google in 2014 for about $500 million. He earned his PhD at the Dalle Molle Institute for Artificial Intelligence in Switzerland and also worked at the Gatsby Computational Neuroscience Unit in London. His dissertation Machine Superintelligence concludes:

If there is ever to be something approaching absolute power, a superintelligent machine would come close. By definition, it would be capable of achieving a vast range of goals in a wide range of environments. If we carefully prepare for this possibility in advance, not only might we avert disaster, we might bring about an age of prosperity unlike anything seen before.

In a later interview, he states:

AI is now where the internet was in 1988. Demand for machine learning skills is quite strong in specialist applications (search companies like Google, hedge funds and bio-informatics) and is growing every year. I expect this to become noticeable in the mainstream around the middle of the next decade. I expect a boom in AI around 2020 followed by a decade of rapid progress, possibly after a market correction. Human level AI will be passed in the mid 2020’s, though many people won’t accept that this has happened. After this point the risks associated with advanced AI will start to become practically important…I don’t know about a “singularity”, but I do expect things to get really crazy at some point after human level AGI has been created. That is, some time from 2025 to 2040.

He and his co-founders Demis Hassabis and Mustafa Suleyman have signed the Future of Life Institute petition on AI risks, and one of their conditions for joining Google was that the company agree to set up an AI Ethics Board to investigate these issues.

Steve Omohundro (wiki) is a former Professor of Computer Science at University of Illinois, founder of the Vision and Learning Group and the Center for Complex Systems Research, and inventor of various important advances in machine learning and machine vision. His work includes lip-reading robots, the StarLisp parallel programming language, and geometric learning algorithms. He currently runs Self-Aware Systems, “a think-tank working to ensure that intelligent technologies are beneficial for humanity”. His paper Basic AI Drives helped launch the field of machine ethics by pointing out that superintelligent systems will converge upon certain potentially dangerous goals. He writes:

We have shown that all advanced AI systems are likely to exhibit a number of basic drives. It is essential that we understand these drives in order to build technology that enables a positive future for humanity. Yudkowsky has called for the creation of ‘friendly AI’. To do this, we must develop the science underlying ‘utility engineering’, which will enable us to design utility functions that will give rise to the consequences we desire…The rapid pace of technological progress suggests that these issues may become of critical importance soon.”

See also his section here on “Rational AI For The Greater Good”.

Murray Shanahan (site) earned his PhD in Computer Science from Cambridge and is now Professor of Cognitive Robotics at Imperial College London. He has published papers in areas including robotics, logic, dynamic systems, computational neuroscience, and philosophy of mind. He is currently writing a book The Technological Singularity which will be published in August; Amazon’s blurb says:

Shanahan describes technological advances in AI, both biologically inspired and engineered from scratch. Once human-level AI — theoretically possible, but difficult to accomplish — has been achieved, he explains, the transition to superintelligent AI could be very rapid. Shanahan considers what the existence of superintelligent machines could mean for such matters as personhood, responsibility, rights, and identity. Some superhuman AI agents might be created to benefit humankind; some might go rogue. (Is Siri the template, or HAL?) The singularity presents both an existential threat to humanity and an existential opportunity for humanity to transcend its limitations. Shanahan makes it clear that we need to imagine both possibilities if we want to bring about the better outcome.

Marcus Hutter (wiki) is a professor in the Research School of Computer Science at Australian National University. He has previously worked with the Dalle Molle Institute for Artificial Intelligence and National ICT Australia, and done work on reinforcement learning, Bayesian sequence prediction, complexity theory, Solomonoff induction, computer vision, and genomic profiling. He has also written extensively on the Singularity. In Can Intelligence Explode?, he writes:

This century may witness a technological explosion of a degree deserving the name singularity. The default scenario is a society of interacting intelligent agents in a virtual world, simulated on computers with hyperbolically increasing computational resources. This is inevitably accompanied by a speed explosion when measured in physical time units, but not necessarily by an intelligence explosion…if the virtual world is inhabited by interacting free agents, evolutionary pressures should breed agents of increasing intelligence that compete about computational resources. The end-point of this intelligence evolution/acceleration (whether it deserves the name singularity or not) could be a society of these maximally intelligent individuals. Some aspect of this singularitarian society might be theoretically studied with current scientific tools. Way before the singularity, even when setting up a virtual society in our imagine, there are likely some immediate difference, for example that the value of an individual life suddenly drops, with drastic consequences.

Jurgen Schmidhuber (wiki) is Professor of Artificial Intelligence at the University of Lugano and former Professor of Cognitive Robotics at the Technische Universitat Munchen. He makes some of the most advanced neural networks in the world, has done further work in evolutionary robotics and complexity theory, and is a fellow of the European Academy of Sciences and Arts. In Singularity Hypotheses, Schmidhuber argues that “if future trends continue, we will face an intelligence explosion within the next few decades”. When asked directly about AI risk on a Reddit AMA thread, he answered:

Stuart Russell’s concerns [about AI risk] seem reasonable. So can we do anything to shape the impacts of artificial intelligence? In an answer hidden deep in a related thread I just pointed out: At first glance, recursive self-improvement through Gödel Machines seems to offer a way of shaping future superintelligences. The self-modifications of Gödel Machines are theoretically optimal in a certain sense. A Gödel Machine will execute only those changes of its own code that are provably good, according to its initial utility function. That is, in the beginning you have a chance of setting it on the “right” path. Others, however, may equip their own Gödel Machines with different utility functions. They will compete. In the resulting ecology of agents, some utility functions will be more compatible with our physical universe than others, and find a niche to survive. More on this in a paper from 2012.

Richard Sutton (wiki) is professor and iCORE chair of computer science at University of Alberta. He is a fellow of the Association for the Advancement of Artificial Intelligence, co-author of the most-used textbook on reinforcement learning, and discoverer of temporal difference learning, one of the most important methods in the field.

In his talk at the Future of Life Institute’s Future of AI Conference, Sutton states that there is “certainly a significant chance within all of our expected lifetimes” that human-level AI will be created, then goes on to say the AIs “will not be under our control”, “will compete and cooperate with us”, and that “if we make superintelligent slaves, then we will have superintelligent adversaries”. He concludes that “We need to set up mechanisms (social, legal, political, cultural) to ensure that this works out well” but that “inevitably, conventional humans will be less important.” He has also mentioned these issues at a presentation to the Gadsby Institute in London and in (of all things) a Glenn Beck book: “Richard Sutton, one of the biggest names in AI, predicts an intelligence explosion near the middle of the century”.

Andrew Davison (site) is Professor of Robot Vision at Imperial College London, leader of the Robot Vision Research Group and Dyson Robotics Laboratory, and inventor of the computerized localization-mapping system MonoSLAM. On his website, he writes:

At the risk of going out on a limb in the proper scientific circles to which I hope I belong(!), since 2006 I have begun to take very seriously the idea of the technological singularity: that exponentially increasing technology might lead to super-human AI and other developments that will change the world utterly in the surprisingly near future (i.e. perhaps the next 20–30 years). As well as from reading books like Kurzweil’s ‘The Singularity is Near’ (which I find sensational but on the whole extremely compelling), this view comes from my own overview of incredible recent progress of science and technology in general and specificially in the fields of computer vision and robotics within which I am personally working. Modern inference, learning and estimation methods based on Bayesian probability theory (see Probability Theory: The Logic of Science or free online version, highly recommended), combined with the exponentially increasing capabilities of cheaply available computer processors, are becoming capable of amazing human-like and super-human feats, particularly in the computer vision domain.

It is hard to even start thinking about all of the implications of this, positive or negative, and here I will just try to state facts and not offer much in the way of opinions (though I should say that I am definitely not in the super-optimistic camp). I strongly think that this is something that scientists and the general public should all be talking about. I’ll make a list here of some ‘singularity indicators’ I come across and try to update it regularly. These are little bits of technology or news that I come across which generally serve to reinforce my view that technology is progressing in an extraordinary, faster and faster way that will have consequences few people are yet really thinking about.

Alan Turing and I. J. Good (wiki, wiki) are men who need no introduction. Turing invented the mathematical foundations of computing and shares his name with Turing machines, Turing completeness, and the Turing Test. Good worked with Turing at Bletchley Park, helped build some of the first computers, and invented various landmark algorithms like the Fast Fourier Transform. In his paper “Can Digital Machines Think?”, Turing writes:

Let us now assume, for the sake of argument, that these machines are a genuine possibility, and look at the consequences of constructing them. To do so would of course meet with great opposition, unless we have advanced greatly in religious tolerance since the days of Galileo. There would be great opposition from the intellectuals who were afraid of being put out of a job. It is probable though that the intellectuals would be mistaken about this. There would be plenty to do in trying to keep one’s intelligence up to the standards set by the machines, for it seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers…At some stage therefore we should have to expect the machines to take control.

During his time at the Atlas Computer Laboratory in the 60s, Good expanded on this idea in Speculations Concerning The First Ultraintelligent Machine, which argued:

Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an ‘intelligence explosion,’ and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make

\* \* \* \* \* \* \* \* \* \*

I worry this list will make it look like there is some sort of big “controversy” in the field between “believers” and “skeptics” with both sides lambasting the other. This has not been my impression.

When I read the articles about skeptics, I see them making two points over and over again. First, we are nowhere near human-level intelligence right now, let alone superintelligence, and there’s no obvious path to get there from here. Second, if you start demanding bans on AI research then you are an idiot.

I agree whole-heartedly with both points. So do the leaders of the AI risk movement.

A survey of AI researchers (Muller & Bostrom, 2014) finds that on average they expect a 50% chance of human-level AI by 2040 and 90% chance of human-level AI by 2075. On average, 75% believe that superintelligence (“machine intelligence that greatly surpasses the performance of every human in most professions”) will follow within thirty years of human-level AI. There are some reasons to worry about sampling bias based on eg people who take the idea of human-level AI seriously being more likely to respond (though see the attempts made to control for such in the survey) but taken seriously it suggests that most AI researchers think there’s a good chance this is something we’ll have to worry about within a generation or two.

But outgoing MIRI director Luke Muehlhauser and Future of Humanity Institute director Nick Bostrom are both on record saying they have significantly later timelines for AI development than the scientists in the survey. If you look at Stuart Armstrong’s AI Timeline Prediction Data there doesn’t seem to be any general law that the estimates from AI risk believers are any earlier than those from AI risk skeptics. In fact, the latest estimate on the entire table is from Armstrong himself; Armstrong nevertheless currently works at the Future of Humanity Institute raising awareness of AI risk and researching superintelligence goal alignment.

The difference between skeptics and believers isn’t about when human-level AI will arrive, it’s about when we should start preparing.

Which brings us to the second non-disagreement. The “skeptic” position seems to be that, although we should probably get a couple of bright people to start working on preliminary aspects of the problem, we shouldn’t panic or start trying to ban AI research.

The “believers”, meanwhile, insist that although we shouldn’t panic or start trying to ban AI research, we should probably get a couple of bright people to start working on preliminary aspects of the problem.

Yann LeCun is probably the most vocal skeptic of AI risk. He was heavily featured in the Popular Science article, was quoted in the Marginal Revolution post, and spoke to KDNuggets and IEEE on “the inevitable singularity questions”, which he describes as “so far out that we can write science fiction about it”. But when asked to clarify his position a little more, he said:

Elon [Musk] is very worried about existential threats to humanity (which is why he is building rockets with the idea of sending humans colonize other planets). Even if the risk of an A.I. uprising is very unlikely and very far in the future, we still need to think about it, design precautionary measures, and establish guidelines. Just like bio-ethics panels were established in the 1970s and 1980s, before genetic engineering was widely used, we need to have A.I.-ethics panels and think about these issues. But, as Yoshua [Bengio] wrote, we have quite a bit of time

Eric Horvitz is another expert often mentioned as a leading voice of skepticism and restraint. His views have been profiled in articles like Out Of Control AI Will Not Kill Us, Believes Microsoft Research Chief and Nothing To Fear From Artificial Intelligence, Says Microsoft’s Eric Horvitz. But here’s what he says in a longer interview with NPR:

KASTE: Horvitz doubts that one of these virtual receptionists could ever lead to something that takes over the world. He says that’s like expecting a kite to evolve into a 747 on its own. So does that mean he thinks the singularity is ridiculous?

Mr. HORVITZ: Well, no. I think there’s been a mix of views, and I have to say that I have mixed feelings myself.

KASTE: In part because of ideas like the singularity, Horvitz and other A.I. scientists have been doing more to look at some of the ethical issues that might arise over the next few years with narrow A.I. systems. They’ve also been asking themselves some more futuristic questions. For instance, how would you go about designing an emergency off switch for a computer that can redesign itself?

Mr. HORVITZ: I do think that the stakes are high enough where even if there was a low, small chance of some of these kinds of scenarios, that it’s worth investing time and effort to be proactive.

Which is pretty much the same position as a lot of the most zealous AI risk proponents. With enemies like these, who needs friends?

A Slate article called Don’t Fear Artificial Intelligence also gets a surprising amount right:

As Musk himself suggests elsewhere in his remarks, the solution to the problem [of AI risk] lies in sober and considered collaboration between scientists and policymakers. However, it is hard to see how talk of “demons” advances this noble goal. In fact, it may actively hinder it.

First, the idea of a Skynet scenario itself has enormous holes. While computer science researchers think Musk’s musings are “not completely crazy,” they are still awfully remote from a world in which AI hype masks less artificially intelligent realities that our nation’s computer scientists grapple with:

Yann LeCun, the head of Facebook’s AI lab, summed it up in a Google+ post back in 2013: “Hype is dangerous to AI. Hype killed AI four times in the last five decades. AI Hype must be stopped.”…LeCun and others are right to fear the consequences of hype. Failure to live up to sci-fi–fueled expectations, after all, often results in harsh cuts to AI research budgets.

AI scientists are all smart people. They have no interest in falling into the usual political traps where they divide into sides that accuse each other of being insane alarmists or ostriches with their heads stuck in the sand. It looks like they’re trying to balance the need to start some preliminary work on a threat that looms way off in the distance versus the risk of engendering so much hype that it starts a giant backlash.

This is not to say that there aren’t very serious differences of opinion in how quickly we need to act. These seem to hinge mostly on whether it’s safe to say “We’ll deal with the problem when we come to it” or whether there will be some kind of “hard takeoff” which will take events out of control so quickly that we’ll want to have done our homework beforehand. I continue to see less evidence than I’d like that most AI researchers with opinions understand the latter possibility, or really any of the technical work in this area. Heck, the Marginal Revolution article quotes an expert as saying that superintelligence isn’t a big risk because “smart computers won’t create their own goals”, even though anyone who has read Bostrom knows that this is exactly the problem.

There is still a lot of work to be done. But cherry-picked articles about how “real AI researchers don’t worry about superintelligence” aren’t it.

[thanks to some people from MIRI and FLI for help with and suggestions on this post]

EDIT: Investigate for possible inclusion: Fredkin, Minsky

# No Time Like The Present For AI Safety Work

I.

On the recent post on AI risk, a commenter challenged me to give the short version of the argument for taking it seriously. I said something like:

1. If humanity doesn’t blow itself up, eventually we will create human-level AI.

2. If humanity creates human-level AI, technological progress will continue and eventually reach far-above-human-level AI

3. If far-above-human-level AI comes into existence, eventually it will so overpower humanity that our existence will depend on its goals being aligned with ours

4. It is possible to do useful research now which will improve our chances of getting the AI goal alignment problem right

5. Given that we can start research now we probably should, since leaving it until there is a clear and present need for it is unwise

I placed very high confidence (>95%) on each of the first three statements – they’re just saying that if trends continue moving towards a certain direction without stopping, eventually they’ll get there. I had lower confidence (around 50%) on the last two statements.

Commenters tended to agree with this assessment; nobody wanted to seriously challenge any of 1-3, but a lot of people said they just didn’t think there was any point in worrying about AI now. We ended up in an extended analogy about illegal computer hacking. It’s a big problem that we’ve never been able to fully address – but if Alan Turing had gotten it into his head to try to solve it in 1945, his ideas might have been along the lines of “Place your punch cards in a locked box where German spies can’t read them.” Wouldn’t trying to solve AI risk in 2015 end in something equally cringeworthy?

Maybe. But I disagree for a couple reasons, some of them broad and meta-level, some of them more focused and object level. The most important meta-level consideration is: if you’re accepting points 1 to 3 – that is, you accept that eventually the human race is going to go extinct or worse if we can’t figure out AI goal alignment – do you really think our chances of making a dent in the problem today are so low that saying “Yes, we’re on a global countdown to certain annhilation, but it would be an inefficient use of resources to even investigate if we could do anything about it at this point”? What is this amazing other use of resources that you prefer? Like, go on and grumble about Pascal’s Wager, but you do realize we just paid Floyd Mayweather ten times more money than has been spent on AI risk total throughout all of human history to participate in a single boxing fight, right?

(if AI boxing got a tenth as much attention, or a hundredth as much money, as AI boxing, the world would be a much safer place)

But I want to make a stronger claim: not just that dealing with AI risk is more important than boxing, but that it is as important as all the other things we consider important, like curing diseases and detecting asteroids and saving the environment. That requires at least a little argument for why progress should indeed be possible at this early stage.

And I think progress is possible insofar as this is a philosophical and not a technical problem. Right now the goal isn’t “write the code that will control the future AI”, it’s “figure out the broad category of problem we have to deal with.” Let me give some examples of open problems to segue into a discussion of why these problems are worth working on now.

II.

Problem 1: Wireheading

Some people have gotten electrodes implanted in their brains for therapeutic or research purposes. When the electrodes are in certain regions, most notably the lateral hypothalamus, the people become obsessed with stimulating them as much as possible. If you give them the stimulation button, they’ll press it thousands of times per hour; if you try to take the stimulation button away from them, they’ll defend it with desperation and ferocity. Their life and focus narrows to a pinpoint, normal goals like love and money and fame and friendship forgotten in the relentless drive to stimulate the electrode as much as possible.

This fits pretty well with what we know of neuroscience. The brain (OVERSIMPLIFICATION WARNING) represents reward as electrical voltage at a couple of reward centers, then does whatever tends to maximize that reward. Normally this works pretty well; when you fulfill a biological drive like food or sex, the reward center responds with little bursts of reinforcement, and so you continue fulfilling your biological drives. But stimulating the reward center directly with an electrode increases it much more than waiting for your brain to send little bursts of stimulation the natural way, so this activity is by definition the most rewarding possible. A person presented with the opportunity to stimulate the reward center directly will forget about all those indirect ways of getting reward like “living a happy life” and just press the button attached to the electrode as much as possible.

This doesn’t even require any brain surgery – drugs like cocaine and meth are addictive in part because they interfere with biochemistry to increase the level of stimulation in reward centers.

And computers can run into the same issue. I can’t find the link, but I do remember hearing about an evolutionary algorithm designed to write code for some application. It generated code semi-randomly, ran it by a “fitness function” that assessed whether it was any good, and the best pieces of code were “bred” with each other, then mutated slightly, until the result was considered adequate.

They ended up, of course, with code that hacked the fitness function and set it to some absurdly high integer.

These aren’t isolated incidents. Any mind that runs off of reinforcement learning with a reward function – and this seems near-universal in biological life-forms and is increasingly common in AI – will have the same design flaw. The main defense against it this far is simple lack of capability: most computer programs aren’t smart enough for “hack your own reward function” to be an option; as for humans, our reward centers are hidden way inside our heads where we can’t get to it. A hypothetical superintelligence won’t have this problem: it will know exactly where its reward center is and be intelligent enough to reach it and reprogram it.

The end result, unless very deliberate steps are taken to prevent it, is that an AI designed to cure cancer hacks its own module determining how much cancer has been cured and sets it to the highest number its memory is capable of representing. Then it goes about acquiring more memory so it can represent higher numbers. If it’s superintelligent, its options for acquiring new memory include “take over all the computing power in the world” and “convert things that aren’t computers into computers.” Human civilization is a thing that isn’t a computer.

This is not some exotic failure mode that a couple of extremely bizarre designs can fall into; this may be the natural course for a sufficiently intelligent reinforcement learner.

Problem 2: Weird Decision Theory

Pascal’s Wager is a famous argument for why you should join an organized religion. Even if you believe God is vanishingly unlikely to exist, the consequence of being wrong (Hell) is so great, and the benefits of being right (not having to go to church on Sundays) so comparatively miniscule, that you should probably just believe in God to be on the safe side. Although there are many objections based on the specific content of religion (does God really want someone to believe based on that kind of analysis?) the problem can be generalized into a form where you can make an agent do anything merely by promising a spectacularly high reward; if the reward is high enough, it will overrule any concerns the agent has about your inability to deliver it.

This is a problem of decision theory which is unrelated to questions of intelligence. A very intelligent person might be able to calculate the probability of God existing very accurately, and they might be able to estimate the exact badness of Hell, but without a good decision theory intelligence alone can’t save you from Pascal’s Wager – in fact, intelligence is what lets you do the formal mathematical calculations telling you to take the bet.

Humans are pretty resistant to this kind of problem – most people aren’t moved by Pascal’s Wager, even if they can’t think of a specific flaw in it – but it’s not obvious how exactly we gain our resistance. Computers, which are infamous for relying on formal math but having no common sense, won’t have that kind of resistance unless it gets built in. And building it in is a really hard problem. Most hacks that eliminate Pascal’s Wager without having a deep understanding of where (or whether) the formal math is going on just open up more loopholes somewhere else. A solution based on a deep understanding of where the formal math goes wrong, and which preserves the power of the math to solve everyday situations, has as far as I know not yet been developed. Worse, once we solve Pascal’s Wager, there are a couple of dozen very similar decision-theoretic paradoxes that may require entirely different solutions.

This is not a cute little philosophical trick. A sufficiently good “hacker” could subvert a galaxy-spanning artificial intelligence just by threatening (with no credibility) to inflict a spectacularly high punishment on it if it didn’t do what the hacker wanted; if the AI wasn’t Pascal-proofed, it would decide to do whatever the hacker said.

Problem 3: The Evil Genie Effect

Everyone knows the problem with computers is that they do what you say rather than what you mean. Nowadays that just means that a program runs differently when you forget a close-parenthesis, or websites show up weird if you put the HTML codes in the wrong order. But it might lead an artificial intelligence to seriously misinterpret natural language orders.

Age of Ultron actually gets this one sort of right. Tony Stark orders his super-robot Ultron to bring peace to the world; Ultron calculates that the fastest and most certain way to bring peace is to destroy all life. As far as I can tell, Ultron is totally 100% correct about this and in some real-world equivalent that is exactly what would happen. We would get pretty much the same effect by telling an AI to “cure cancer” or “end world hunger” or any of a thousand other things.

I promise the gender gap, among many other things, will be eliminated. https://t.co/9KDLvaNFob

— Sweet Meteor O'Death (@smod2016) May 24, 2015

Even Isaac Asimov’s Three Laws of Robotics would take about thirty seconds to become horrible abominations. The First Laws says a robot cannot harm a human being or allow through inaction a human being to come to harm. “Not taking over the government and banning cigarettes” counts as allowing through inaction a human being to come to harm. So does “not locking every human in perfectly safe stasis fields for all eternity.”

There is no way to compose an order specific enough to explain exactly what we mean by “do not allow through inaction a human to come to harm” – go ahead, try it – unless the robot is already willing to do what we mean, rather than what we say. This is not a deal-breaker, since AIs may indeed by smart enough to understand what we mean, but our desire that they do so will have to be programmed into them directly, from the ground up. Part of SIAI’s old vision of “causal validity semantics” seems to be about laying a groundwork for this program.

But this just leads to a second problem: we don’t always know what we mean by something. The question of “how do we balance the ethical injunction to keep people safe with the ethical injunction to preserve human freedom?” is a pretty hot topic in politics right now, presenting itself in everything from gun control to banning Big Gulp cups. It seems to involve balancing out everything we value – how important are Big Gulp cups to us, anyway? – and combining cost-benefit calculations with sacred principles. Any AI that couldn’t navigate that moral labyrinth might end up ending world hunger by killing all starving people, or refusing else to end world hunger by inventing new crops because the pesticides for them might kill an insect.

But the more you try to study ethics, the more you realize they’re really really complicated and so far resist simplification to the sort of formal system that a computer has any hope of understanding. Utilitarianism is almost computer-readable, but it runs into various paradoxes at the edges, and even without those you’d need to have a set of utility weights for everything in the world.

This is a problem we have yet to solve with humans – most of the humans in the world have values that we consider abhorrent, and accept tradeoffs we consider losing propositions. Dealing with an AI whose mind is no more different to mine than that of fellow human being Pat Robertson would from my perspective be a clear-cut case of failure.

[EDIT: I’m told I’m not explaining this very well. This might be better.]

III.

My point in raising these problems wasn’t to dazzle anybody with interesting philosophical issues. It’s to prove a couple of points:

First, there are some very basic problems that affect broad categories of minds, like “all reinforcement learners” or “all minds that make decisions with formal math”. People often speculate that at this early stage we can’t know anything about the design of future AIs. But I would find it extraordinarily surprising if they used neither reinforcement learning or formal mathematical decision-making.

Second, these problems aren’t obvious to most people. These are weird philosophical quandaries, not things that are obvious to everybody with even a little bit of domain knowledge.

Third, these problems have in fact been thought of. Somebody, whether it was a philosopher or a mathematician or a neuroscientist, sat down and thought “Hey, wait, reinforcement learners are naturally vulnerable to wireheading, which would explain why this same behavior shows up in all of these different domains.”

Fourth, these problems suggest research programs that can be pursued right now, at least in a preliminary way. Why do humans resist Pascal’s Wager so effectively? Can our behavior in high-utility, low-probability situations be fitted to a function that allows a computer to make the same decisions we do? What are the best solutions to the related decision theory problems? How come a human can understand the concept of wireheading, yet not feel any compulsion to seek a brain electrode to wirehead themselves with? Is there a way to design a mind that could wirehead a few times, feel and understand the exact sensation, and yet feel no compulsion to wirehead further? How could we create an idea of human ethics and priorities formal enough to stick into a computer?

I think when people hear “we should start, right now in 2015, working on AI goal alignment issues” they think that somebody wants to write a program that can be imported directly into a 2075 AI to provide it with an artificial conscience. Then they think “No way you can do something that difficult this early on.”

But that isn’t what anybody’s proposing. What we’re proposing is to get ourselves acquainted with the general philosophical problems that affect a broad subset of minds, then pursue the neuroscientific, mathematical, and philosophical investigations necessary to have a good understanding of them by the time the engineering problem comes up.

By analogy, we are nowhere near having spaceships that can travel at even half the speed of light. But we already know the biggest obstacle that an FTL spaceship is going to face (relativity and the light-speed limit) and we already have some ideas for getting around it (the Alcubierre drive). We can’t come anywhere close to building an Alcubierre drive. But if we discover how to make near-lightspeed spaceships in 2100, and for some reason the fate of Earth depends on having faster-than-light spaceships by 2120, it’ll probably be nice that we did all of our Theory-Of-Relativity-discovering early so that we’re not wasting half that time interval debating basic physics.

The question “Can we do basic AI safety research now?” is silly because we have already done some basic AI safety research successfully. It’s led to understanding issues like the three problems mentioned above, and many more. There are even a couple of answers now, although they’re at technical levels much lower than any of those big questions. Every step we finish now is one that we don’t have to waste valuable time retracing during the crunch period.

IV.

That last section discussed my claim 4, that there’s research we can do now that will help. That leaves claim 5 – given that we can do research now, we should, because we can’t just trust our descendents in the crunch time to sort things out on their own without our help, using their better model of what eventual AI might look like. There are a couple of reasons for this

Reason 1: The Treacherous Turn

Our descendents’ better models of AI might be actively misleading. Things that work for subhuman or human level intelligences might fail for superhuman intelligences. Empirical testing won’t be able to figure this out without help from armchair philosophy.

Pity poor evolution. It had hundreds of millions of years to evolve defenses against heroin – which by the way affects rats much as it does humans – but it never bothered. Why not? Because until the past century, there wasn’t anything around intelligent enough to synthesize pure heroin. So heroin addiction just wasn’t a problem anything had to evolve to deal with. A brain design that looks pretty good in stupid animals like rats and cows becomes very dangerous when put in the hands (well, heads) of humans smart enough to synthesize heroin or wirehead their own pleasure centers.

The same is true of AI. Dog-level AIs aren’t going to learn to hack their own reward mechanism. Even human level AIs might not be able to – I couldn’t hack a robot reward mechanism if it were presented to me. Superintelligences can. What we might see is reinforcement-learning AIs that work very well at the dog level, very well at the human level, then suddenly blow up at the superhuman level, by which it’s time it’s too late to stop them.

This is a common feature of AI safety failure modes. If you tell me, as a mere human being, to “make peace”, then my best bet might be to become Secretary-General of the United Nations and learn to negotiate very well. Arm me with a few thousand nukes, and it’s a different story. A human-level AI might pursue its peace-making or cancer-curing or not-allowing-human-harm-through-inaction-ing through the same prosocial avenues as humans, then suddenly change once it became superintelligent and new options became open. Indeed, the point that will activate the shift is precisely that no humans are able to stop it. If humans can easily shut an AI down, then the most effective means of curing cancer will be for it to research new medicines (which humans will support); if humans can no longer stop an AI, the most effective means of curing cancer is destroying humanity (since it will no longer matter that humans will fight back).

In his book, Nick Bostrom calls this pattern “the treacherous turn”, and it will doom anybody who plans to just wait until the AIs exist and then solve their moral failings through trial and error and observation. The better plan is to have a good philosophical understanding of exactly what’s going on, so we can predict these turns ahead of time and design systems that avoid them from the ground up.

Reason 2: Hard Takeoff

Nathan Taylor of Praxtime writes:

Arguably most of the current “debates” about AI Risk are mere proxies for a single, more fundamental disagreement: hard versus soft takeoff.

Soft takeoff means AI progress takes a leisurely course from the subhuman level to the dumb-human level to the smarter-human level to the superhuman level over many decades. Hard takeoff means the same course takes much shorter, maybe days to months.

It seems in theory that by hooking a human-level AI to a calculator app, we can get it to the level of a human with lightning-fast calculation abilities. By hooking it up to Wikipedia, we can give it all human knowledge. By hooking it up to a couple extra gigabytes of storage, we can give it photographic memory. By giving it a few more processors, we can make it run a hundred times faster, such that a problem that takes a normal human a whole day to solve only takes the human-level AI fifteen minutes.

So we’ve already gone from “mere human intelligence” to “human with all knowledge, photographic memory, lightning calculations, and solves problems a hundred times faster than anyone else.” This suggests that “merely human level intelligence” isn’t mere.

The next problem is “recursive self-improvement”. Maybe this human-level AI armed with photographic memory and a hundred-time-speedup takes up computer science. Maybe, with its ability to import entire textbooks in seconds, it becomes very good at computer science. This would allow it to fix its own algorithms to make itself even more intelligent, which would allow it to see new ways to make itself even more intelligent, and so on. The end result is that it either reaches some natural plateau or becomes superintelligent in the blink of an eye.

If it’s the second one, “wait for the first human-level intelligences and then test them exhaustively” isn’t going to cut it. The first human-level intelligence will become the first superintelligence too quickly to solve even the first of the hundreds of problems involved in machine goal-alignment.

And although I haven’t seen anyone else bring this up, I’d argue that even the hard-takeoff scenario might be underestimating the risks.

Imagine that for some reason having two hundred eyes is the killer app for evolution. A hundred ninety-nine eyes are useless, no better than the usual two, but once you get two hundred, your species dominates the world forever.

The really hard part of having two hundred eyes is evolving the eye at all. After you’ve done that, having two hundred of them is very easy. But it might be that it would take eons and eons before any organism reached the two hundred eye sweet spot. Having dozens of eyes is such a useless waste of energy that evolution might never get to the point where it could test the two-hundred-eyed design.

Consider that the same might be true for intelligence. The hard part is evolving so much as a tiny rat brain. Once you’ve got that, getting a human brain, with its world-dominating capabilities, is just a matter of scaling up. But since brains are metabolically wasteful and not that useful before the technology-discovering point, it took eons before evolution got there.

There’s a lot of evidence that this is true. First of all, humans evolved from chimps in just a couple of million years. That’s too short to redesign the mind from the ground up, or even invent any interesting new evolutionary “technologies”. It’s just enough time for evolution to alter the scale and add a couple of efficiency tweaks. But monkeys and apes were around for tens of millions of years before evolution bothered.

Second, dolphins are almost as intelligent as humans. But they last shared a common ancestor with us something like fifty million years ago. Either humans and dolphins both evolved fifty million years worth of intelligence “technologies” independently of each other, or else the most recent common ancestor had most of what was necessary for intelligence and humans and dolphins were just the two animals in that vast family tree for whom using them to their full extent became useful. But the most recent common ancestor of humans and dolphins was probably not much more intelligent than a rat itself.

Third, humans can gain intelligence frighteningly quickly when the evolutionary pressures are added. If Cochran is right, Ashkenazi gained ten IQ points in a thousand years. Torsion dystonia sufferers can gain five or ten IQ points from a single mutation. All of this suggests a picture where intelligence is easy to change, but evolution has decided it just isn’t worth it except in very specific situations.

If this is right, then the first rat-level AI will contain most of the interesting discoveries needed to build the first human-level AI and the first superintelligent AI. People tend to say things like “Well, we might have AI as smart as a rat soon, but it will be a long time after that before they’re anywhere near human-level”. But that’s assuming you can’t turn the rat into the human just by adding more processing power or more simulated neurons or more connections or whatever. Anything done on a computer doesn’t need to worry about metabolic restrictions.

Reason 3: Everyday Ordinary Time Constraints

Bostrom and Mueller surveyed AI researchers about when they expected human-level AI. The median date was 2040. That’s 25 years.

People have been thinking about Pascal’s Wager (for example) for 345 years now without coming up with any fully generalizable solutions. If that turns out to be a problem for AI, we have 25 more years to solve not only the Wager, but the entire class of problems to which it belongs. Even barring scenarios like unexpected hard takeoffs or treacherous turns, and accepting that if we can solve the problem in 25 years everything will be great, that’s not a lot of time.

During the 1956 Dartmouth Conference on AI, top researchers made a plan toward reaching human-level artificial intelligence, and gave themselves two months to teach computers to understand human language. In retrospect, this might have been mildly optimistic.

But now machine translation is a thing, people are making some good progress in some of the hard problems – and when people bring up problems like decision theory, or wireheading, or goal alignment, people just say “Oh, we have plenty of time”.

But expecting to solve those problems in a few years might be just as optimistic as expecting to solve machine language translation in two months. Sometimes problems are harder than you think, and it’s worth starting on them early just in case.

All of this means it’s well worth starting armchair work on AI safety now. I won’t say the entire resources of our civilization need to be sunk into it immediately, and I’ve ever heard some people in the field say that after Musk’s $10 million donation money is no longer the most important bottleneck to advancing these ideas. I’m not even sure public exposure is a bottleneck anymore; the median person who watches a movie about killer robots is probably doing more harm than good. If the bottleneck is anything at all, it’s probably intelligent people in relevant fields – philosophy, AI, math, and neuroscience – applying brainpower to these issues and encouraging their colleagues to take them seriously.

# That Chocolate Study

Several of you asked me to write about that chocolate article that went viral recently. From I Fooled Millions Into Thinking Chocolate Helps Weight Loss. Here’s How:

“Slim by Chocolate!” the headlines blared. A team of German researchers had found that people on a low-carb diet lost weight 10 percent faster if they ate a chocolate bar every day. It made the front page of Bild, Europe’s largest daily newspaper, just beneath their update about the Germanwings crash. From there, it ricocheted around the internet and beyond, making news in more than 20 countries and half a dozen languages. It was discussed on television news shows. It appeared in glossy print, most recently in the June issue of Shape magazine (“Why You Must Eat Chocolate Daily,” page 128). Not only does chocolate accelerate weight loss, the study found, but it leads to healthier cholesterol levels and overall increased well-being. The Bild story quotes the study’s lead author, Johannes Bohannon, Ph.D., research director of the Institute of Diet and Health: “The best part is you can buy chocolate everywhere.”

I am Johannes Bohannon, Ph.D. Well, actually my name is John, and I’m a journalist. I do have a Ph.D., but it’s in the molecular biology of bacteria, not humans. The Institute of Diet and Health? That’s nothing more than a website.

Other than those fibs, the study was 100 percent authentic. My colleagues and I recruited actual human subjects in Germany. We ran an actual clinical trial, with subjects randomly assigned to different diet regimes. And the statistically significant benefits of chocolate that we reported are based on the actual data. It was, in fact, a fairly typical study for the field of diet research. Which is to say: It was terrible science. The results are meaningless, and the health claims that the media blasted out to millions of people around the world are utterly unfounded.

Bohannon goes on to explain that as part of a documentary about “the junk-science diet industry”, he and some collaborators designed a fake study to see if they could convince journalists. They chose to make it about chocolate:

Gunter Frank, a general practitioner in on the prank, ran the clinical trial. Onneken had pulled him in after reading a popular book Frank wrote railing against dietary pseudoscience. Testing bitter chocolate as a dietary supplement was his idea. When I asked him why, Frank said it was a favorite of the “whole food” fanatics. “Bitter chocolate tastes bad, therefore it must be good for you,” he said. “It’s like a religion.”

They recruited 16 (!) participants and divided them into three groups. One group ate their normal diet. Another ate a low-carb diet. And a third ate a low-carb diet plus some chocolate. Both the low-carb group and the low-carb + chocolate group lost weight compared to the control group, but the low-carb + chocolate group lost weight “ten percent faster”, and the difference was “statistically significant”. They also had “better cholesterol readings” and “higher scores on the well-being survey”.

Bohannon admits exactly how he managed this seemingly impressive result – he measured eighteen different parameters (weight, cholesterol, sodium, protein, etc) which virtually guarantees that one will be statistically significant. That one turned out to be weight loss. If it had been sodium, he would have published the study as “Chocolate Lowers Sodium Levels”.

Then he pitched it to various fake for-profit journals until one of them bit. Then he put out a PR release to various media outlets, and they ate it up. They ended up in a bunch of English and German language media including Bild, the Daily Star, Times of India, Cosmopolitan, Irish Examiner, and the Huffington Post.

The people I’ve seen discussing this seem to have drawn five conclusions, four of which are wrong:

Conclusion 1: Haha, I can’t believe people were so gullible that they actually thought chocolate caused weight loss!

Bohannon himself endorses this one, saying bitter chocolate was a favorite of “whole food fanatics” because “Bitter chocolate tastes bad, therefore it must be good for you” and “it’s like a religion.

But actually, there’s lots of previous research supporting health benefits from bitter chocolate, none of which Bohannon seems to be aware of.

A meta-analysis of 42 randomized controlled trials totaling 1297 participants in the American Journal of Clinical Nutrition found that chocolate improved blood pressure, flow-mediated dilatation (a measure of vascular health), and insulin resistance (related to weight gain).

A different meta-analysis of 24 randomized controlled trials totalling 1106 people in the Journal of Nutrition also found that chocolate improved blood pressure, flow-mediated dilatation, and insulin resistance.

A Cochrane Review of 20 randomized controlled trials of 856 people found that chocolate improved blood pressure (it didn’t test for flow-mediated dilatation or insulin resistance)

A study on mice found that mice fed more chocolate flavanols were less likely to gain weight.

An epidemiological study of 1018 people in the United States found an association between frequent chocolate consumption and lower BMI, p < 0.01.  
  
A second epidemiological study of 1458 people in Europe found the same thing, again p < 0.01.  
  
A cohort study of 470 elderly men found chocolate intake was inversely associated with blood pressure and cardiovascular mortality, p less than 0.001, not confounded by the usual suspects.

I wouldn’t find any of these studies alone very convincing. But together, they compensate for each other’s flaws and build a pretty robust structure. So the next flawed conclusion is:

Conclusion 2: This proves that nutrition isn’t a real science and we should all just be in a state of radical skepticism about these things

What we would like to do is a perfect study where we get thousands of people, randomize them to eat-lots-of-chocolate or eat-little-chocolate at birth, then follow their weights over their entire lives. That way we could have a large sample size, perfect randomization, life-long followup, and clear applicability to other people. But for practical and ethical reasons, we can’t do that. So we do a bunch of smaller studies that each capture a few of the features of the perfect study.

First we do animal studies, which can have large sample sizes, perfect randomization, and life-long followup, but it’s not clear whether it applies to humans.

Then we do short randomized controlled trials, which can have large sample sizes, perfect randomization, and human applicability, but which only last a couple of months.

Then we do epidemiological studies, which can have large sample sizes, human applicability, and last for many decades, but which aren’t randomized very well and might be subject to confounders.

This is what happened in the chocolate studies above. Mice fed a strict diet plus chocolate for a long time gain less weight than mice fed the strict diet alone. This is suggestive, but we don’t know if it applies to humans. So we find that in randomized controlled trials, chocolate helps with some proxies for weight gain like insulin resistance. This is even more suggestive, but we don’t know if it lasts. So we find that in epidemiological studies, lifetime chocolate consumption is associated with lifetime good health outcomes. This on its own is suggestive but potentially confounded, but when we combine them with all of the others, they become more convincing.

(am I cheating by combining blood pressure and BMI data? Sort of, but the two measures are correlated)

When all of these paint the same picture, then we start thinking that maybe it’s because our hypothesis is true. Yes, maybe the mouse studies could be related to a feature of mice that doesn’t generalize to humans, and the randomized controlled trial results wouldn’t hold up after a couple of years, and the epidemiological studies are confounded. But that would be extraordinarily bad luck. More likely they’re all getting the same result because they’re all tapping into the same underlying reality.

This is the way science usually works, it’s the way nutrition science usually works, and it’s the way the science of whether chocolate causes weight gain usually works. These are not horrible corrupt disciplines made up entirely of shrieking weight-loss-pill peddlers trying to hawk their wares. They only turn into that when the media takes a single terrible study totally out of context and misrepresents the field.

Conclusion 3: Studies Always Need To Have High Sample Sizes

Here’s another good chocolate-related study: Short-term administration of dark chocolate is followed by a significant increase in insulin sensitivity and a decrease in blood pressure in healthy persons.

Bohannon says:

Our study was doomed by the tiny number of subjects, which amplifies the effects of uncontrolled factors…Which is why you need to use a large number of people, and balance age and gender across treatment group

But I say “Short-term administration…” is a good study despite having an n = 15, one less than the Bohannon study. Why? Well, their procedure was pretty involved, and you wouldn’t be able to get a thousand people to go through the whole rigamarole. On the other hand, their insulin resistance measure thing was nearly twice as high in the dark chocolate group as the white chocolate group, and p < 0.001.  
  
(Another low sample size study that was nevertheless very good: psychiatrists knew that consuming dietary tyramine when taking a MAOI antidepressant can cause a life-threatening hypertensive crisis, but they didn't know how much tyramine it took. In order to find out, they took a dozen people, put them on MAOIs, and then gradually fed them more and more tyramine with doctors standing by to treat the crisis as soon as it started. They found about how much tyramine it took and declared the experiment a success. If the tyramine levels were about the same in all twelve patients, then adding a thousand more patients wouldn’t help much, and it would definitely increase the risk.)

Sample size is important when you’re trying to detect a small effect in the middle of a large amount of natural variation. When you’re looking for a large effect in the middle of no natural variation, sample size doesn’t matter as much. For example, if there was a medicine that would help amputees grow their hands back, I would accept success with a single patient (if it worked) as proof of effectiveness (I suppose I couldn’t be sure it would always work until more patients had been tried, but a single patient would certainly pique my interest). You’re not going after sample size so much as after p-value.

Conclusion 4: P-Values Are Stupid And We Need To Get Rid Of Them

Bohannon says that:

If you measure a large number of things about a small number of people, you are almost guaranteed to get a “statistically significant” result…the letter p seems to have totemic power, but it’s just a way to gauge the signal-to-noise ratio in the data…scientists are getting wise to these problems. Some journals are trying to phase out p value significance testing altogether to nudge scientists into better habits.

Okay, take the “Short-term administration” study above. I would like to be able to say that since it has p < 0.001, we know it's significant. But suppose we're not allowed to do p-values. All I do is tell you "Yeah, there was a study with fifteen people that found chocolate helped with insulin resistance" and you laugh in my face.  
  
Effect size is supposed to help with that. But suppose I tell you "There was a study with fifteen people that found chocolate helped with insulin resistance. The effect size was 0.6." I don't have any intuition at all for whether or not that's consistent with random noise. Do you?  
  
Okay, then they say we’re supposed to report confidence intervals. The effect size was 0.6, with 95% confidence interval of [0.2, 1.0]. Okay. So I check the lower bound of the confidence interval, I see it’s different from zero. But now I’m not transcending the p-value. I’m just using the p-value by doing a sort of kludgy calculation of it myself – “95% confidence interval does not include zero” is the same as “p value is less than 0.05”.

(Imagine that, although I know the 95% confidence interval doesn’t include zero, I start wondering if the 99% confidence interval does. If only there were some statistic that would give me this information!)

But wouldn’t getting rid of p-values prevent “p-hacking”? Maybe, but it would just give way to “d-hacking”. You don’t think you could test for twenty different metabolic parameters and only report the one with the highest effect size? The only difference would be that p-hacking is completely transparent – if you do twenty tests and report a p of 0.05, I know you’re an idiot – but d-hacking would be inscrutable. If you do twenty tests and report that one of them got a d = 0.6, is that impressive? No better than chance? I have no idea. I bet there’s some calculation I could do to find out, but I also bet that it would be a lot harder than just multiplying the value by the number of tests and seeing what happens. [EDIT: On reflection not sure this is true; the possibility of p-hacking is inherent to p-values, but the possibility of d-hacking isn’t inherent to effect size. I don’t actually know how much this would matter in the real world.]

But wouldn’t switching from p-values to effect sizes prevent people from making a big deal about tiny effects that are nevertheless statistically significant? Yes, but sometimes we want to make a big deal about tiny effects that are nevertheless statistically significant! Suppose that Coca-Cola is testing a new product additive, and finds in large epidemiological studies that it causes one extra death per hundred thousand people per year. That’s an effect size of approximately zero, but it might still be statistically significant. And since about a billion people worldwide drink Coke each year, that’s a ten thousand deaths. If Coke said “Nope, effect size too small, not worth thinking about”, they would kill almost two milli-Hitlers worth of people.

Yeah, sure, you can never use p-values again, and run into all of these other problems. Or you can do a Bonferroni correction, which is a very simple adjustment to p-values which corrects for p-hacking. Or instead of taking one study at face value LIKE AN IDIOT you can wait to see if other studies replicate the findings. Remember, the whole point of p-hacking is choosing at random form a bunch of different outcomes, so if two trials both try to p-hack, they’ll end up with different outcomes and the game will be up. Seriously, STOP TRYING TO BASE CONCLUSIONS ON ONE STUDY.

Conclusion 5: Trust Science Journalism Less

This is the one that’s correct.

But it’s not totally correct. Bohannon boasts of getting his findings in a couple of daily newspapers and the Huffington Post. That’s not exactly the cream of the crop. The Economist usually has excellent science journalism. Magazines like Scientific American and Discover can be okay, although even they get hyped. Reddit’s r/science is good, assuming you make sure to always check the comments. And there are individual blogs like Mind the Brain run by researchers in the field that can usually be trusted near-absolutely. Cochrane Collaboration will always have among the best analyses on everything.

If you really want to know what’s going on and can’t be bothered to ferret out all of the brilliant specialists, my highest recommendation goes to Wikipedia. It isn’t perfect, but compared to anything you’d find on a major news site, it’s like night and day. Wikipedia’s Health Effects Of Chocolate page is pretty impressive and backs everything it says up with good meta-analyses and studies in the best journals. Its sentence on the cardiovasuclar effects links to this letter, which is very good.

Do you know why you can trust Wikipedia better than news sites? Because Wikipedia doesn’t obsess over the single most recent study. Are you starting to notice a theme?

For me, the takeaway from this affair is that there is no one-size-fits-all solution to make statistics impossible to hack. Getting rid of p-values is appropriate sometimes, but not other times. Demanding large sample sizes is appropriate sometimes, but not other times. Not trusting silly conclusions like “chocolate causes weight loss” works sometimes but not other times. At the end of the day, you have to actually know what you’re doing. Also, try to read more than one study.

# …And I Show You How Deep The Rabbit Hole Goes

.

Seen on Tumblr, along with associated discussion:



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 nest 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 manuevers 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.”

“Pleeeeeease?” 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.

“Pweeeeeeease?” says Pink. She actually says pweeeeeeease.

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.

# Against Tulip Subsidies

I.

Imagine a little kingdom with a quaint custom: when a man likes a woman, he offers her a tulip; if she accepts, they are married shortly thereafter. A couple who marries sans tulip is considered to be living in sin; no other form of proposal is appropriate or accepted.

One day, a Dutch trader comes to the little kingdom. He explains that his homeland also has a quaint custom involving tulips: they speculate on them, bidding the price up to stratospheric levels. Why, in the Netherlands, a tulip can go for ten times more than the average worker earns in a year! The trader is pleased to find a new source of bulbs, and offers the people of the kingdom a few guilders per tulip, which they happily accept.

Soon other Dutch traders show up and start a bidding war. The price of tulips goes up, and up, and up; first dozens of guilders, then hundreds. Tulip-growers make a fortune, but everyone else is less pleased. Suitors wishing to give a token of their love find themselves having to invest their entire life savings – with no guarantee that the woman will even say yes! Soon, some of the poorest people are locked out of marriage and family-raising entirely.

Some of the members of Parliament are outraged. Marriage is, they say, a human right, and to see it forcibly denied the poor by foreign speculators is nothing less than an abomination. They demand that the King provide every man enough money to guarantee he can buy a tulip. Some objections are raised: won’t it deplete the Treasury? Are we obligated to buy everyone a beautiful flawless bulb, or just the sickliest, grungiest plant that will technically satisfy the requirements of the ritual? If some man continuously proposes to women who reject him, are we obligated to pay for a new bulb each time, subsidizing his stupidity?

The pro-subsidy faction declares that the people asking these question are well-off, and can probably afford tulips of their own, and so from their place of privilege they are trying to raise pointless objections to other people being able to obtain the connubial happiness they themselves enjoy. After the doubters are tarred and feathered and thrown in the river, Parliament votes that the public purse pay for as many tulips as the poor need, whatever the price.

A few years later, another Dutch trader comes to the little kingdom. Everyone asks if he is there to buy tulips, and he says no, the Netherlands’ tulip bubble has long since collapsed, and the price is down to a guilder or two. The people of the kingdom are very surprised to hear that, since the price of their own tulips has never stopped going up, and is now in the range of tens of thousands of guilders. Nevertheless, they are glad that, however high tulip prices may be for them, they know the government is always there to help. Sure, the roads are falling apart and the army is going hungry for lack of rations, but at least everyone who wants to marry is able to do so.

Meanwhile, across the river is another little kingdom that had the same tulip-related marriage custom. They also had a crisis when the Dutch merchants started making the prices go up. But they didn’t have enough money to afford universal tulip subsidies. It was pretty touch-and-go for a while, and a lot of poor people were very unhappy.

But nowadays they use daffodils to mark engagements, and their economy has never been better.

II.

In America, aspiring doctors do four years of undergrad in whatever area they want (I did Philosophy), then four more years of medical school, for a total of eight years post-high school education. In Ireland, aspiring doctors go straight from high school to medical school and finish after five years.

I’ve done medicine in both America and Ireland. The doctors in both countries are about equally good. When Irish doctors take the American standardized tests, they usually do pretty well. Ireland is one of the approximately 100% of First World countries that gets better health outcomes than the United States. There’s no evidence whatsoever that American doctors gain anything from those three extra years of undergrad. And why would they? Why is having a philosophy degree under my belt supposed to make me any better at medicine?

(I guess I might have acquired a talent for colorectal surgery through long practice pulling things out of my ass, but it hardly seems worth it.)

I’ll make another confession. Ireland’s medical school is five years as opposed to America’s four because the Irish spend their first year teaching the basic sciences – biology, organic chemistry, physics, calculus. When I applied to medical school in Ireland, they offered me an accelerated four year program on the grounds that I had surely gotten all of those in my American undergraduate work. I hadn’t. I read some books about them over the summer and did just fine.

Americans take eight years to become doctors. Irishmen can do it in four, and achieve the same result. Each year of higher education at a good school – let’s say an Ivy, doctors don’t study at Podunk Community College – costs about $50,000. So American medical students are paying an extra $200,000 for…what?

Remember, a modest amount of the current health care crisis is caused by doctors’ crippling level of debt. Socially responsible doctors often consider less lucrative careers helping the needy, right up until the bill comes due from their education and they realize they have to make a lot of money right now. We took one look at that problem and said “You know, let’s make doctors pay an extra $200,000 for no reason.”

And to paraphrase Dirkson, $200,000 here, $200,000 there, and pretty soon it adds up to real money. 20,000 doctors graduate in the United States each year; that means the total yearly cost of requiring doctors to have undergraduate degrees is $4 billion. That’s most of the amount of money you’d need to house every homeless person in the country ($10,000 to house one homeless x 600,000 homeless).

I want to be able to say people have noticed the Irish/American discrepancy and are thinking hard about it. I can say that. Just not in the way I would like. Many of the elder doctors I talked to in Ireland wanted to switch to the American system. Not because they thought it would give them better doctors. Just because they said it was more fun working with medical students like myself who were older and a little wiser. The Irish medical students were just out of high school and hard to relate to – us foreigners were four years older than that and had one or another undergraduate subject under our belts. One of my attendings said that it was nice having me around because I’d studied Philosophy in college and that gave our team a touch of class. A touch of class!

This is why, despite my reservations about libertarianism, it’s not-libertarianism that really scares me. Whenever some people without skin in the game are allowed to make decisions for other people, you end up with a bunch of elderly doctors getting together, think “Yeah, things do seem a little classier around here if we make people who are not us pay $200,000, make it so,” and then there goes the money that should have housed all the homeless people in the country.

But more important, it also destroyed my last shred of hope that the current mania for requiring college degrees for everything had a good reason behind it.

III.

The only reason I’m picking on medicine is that it’s so clear. You have your experimental group in the United States, your control group in Ireland, you can see the lack of difference. You can take an American doctor and an Irish doctor, watch them prescribe the same medication in the same situation, and have a visceral feel for “Wait, we just spent $200,000 for no reason.”

But it’s not just medicine. Let me tell you about my family.

There’s my cousin. He wants to be a firefighter. He’s wanted to be a firefighter ever since he was young, and he’s done volunteer work for his local fire department, who have promised him a job. But in order to get it, he has to go do four years of college. You can’t be a firefighter without a college degree. That would be ridiculous. Back in the old days, when people were allowed to become firefighters after getting only thirteen measly years of book learning, I have it on good authority that several major states burnt to the ground.

My mother is a Spanish teacher. After twenty years teaching, with excellent reviews by her students, she pursued a Masters’ in Education because her school was going to pay her more money if she had it. She told me that her professors were incompetent, had never actually taught real students, and spent the entire course pushing whatever was the latest educational fad; however, after paying them thousands of dollars, she got the degree and her school dutifully increased her salary. She is lucky. In several states, teachers are required by law to pursue a Masters’ degree to be allowed to continue teaching. Oddly enough, these states have no better student outcomes than states without this requirement, but this does not seem to affect their zeal for this requirement. Even though many rigorous well-controlled studies have found that presence of absence of a Masters’ degree explains approximately zero percent of variance in teacher quality, many states continue to require it if you want to keep your license, and almost every state will pay you more for having it.

Before taking my current job, I taught English in Japan. I had no Japanese language experience and no teaching experience, but the company I interviewed with asked if I had an undergraduate degree in some subject or other, and that was good enough for them. Meanwhile, I knew people who were fluent in Japanese and who had high-level TOEFL certification. They did not have a college degree so they were not considered.

My ex-girlfriend majored in Gender Studies, but it turned out all of the high-paying gender factories had relocated to China. They solved this problem by going to App Academy, a three month long, $15,000 course that taught programming. App Academy graduates compete for the same jobs as people who have taken computer science in college, a four year long, $200,000 undertaking.

I see no reason to think my family and friends are unique. The overall picture seems to be one of people paying hundreds of thousands of dollars to get a degree in Art History to pursue a job in Sales, or a degree in Spanish Literature to get a job as a middle manager. Or not paying hundreds of thousands of dollars, if they happen to be poor, and so being permanently locked out of jobs as a firefighter or salesman.

IV.

So presidential candidate Bernie Sanders has proposed universal free college tuition.

On the one hand, I sympathize with his goals. If you can’t get any job better than ‘fast food worker’ without a college degree, and poor people can’t afford college degrees, that’s a pretty grim situation, and obviously unfair to the poor.

On the other hand, if can’t you get married without a tulip, and poor people can’t afford tulips, that’s also a pretty grim situation, and obviously unfair to the poor.

But the solution isn’t universal tulip subsidies.

Higher education is in a bubble much like the old tulip bubble. In the past forty years, the price of college has dectupled (quadrupled when adjusting for inflation). It used to be easy to pay for college with a summer job; now it is impossible. At the same time, the unemployment rate of people without college degrees is twice that of people who have them. Things are clearly very bad and Senator Sanders is right to be concerned.

But, well, when we require doctors to get a college degree before they can go to medical school, we’re throwing out a mere $5 billion, barely enough to house all the homeless people in the country. But Senator Sanders admits that his plan would cost $70 billion per year. That’s about the size of the entire economy of Hawaii. It’s enough to give $2000 every year to every American in poverty.

At what point do we say “Actually, no, let’s not do that, and just let people hold basic jobs even if they don’t cough up a a hundred thousand dollars from somewhere to get a degree in Medieval History”?

I’m afraid that Sanders’ plan is a lot like the tulip subsidy idea that started off this post. It would subsidize the continuation of a useless tradition that has turned into a speculation bubble, prevent the bubble from ever popping, and disincentivize people from figuring out a way to route around the problem, eg replacing the tulips with daffodils.

(yes, it is nice to have college for non-economic reasons too, but let’s be honest – if there were no such institution as college, would you, totally for non-economic reasons, suggest the government pay poor people $100,000 to get a degree in Medieval History? Also, anything not related to job-getting can be done three times as quickly by just reading a book.)

If I were Sanders, I’d propose a different strategy. Make “college degree” a protected characteristic, like race and religion and sexuality. If you’re not allowed to ask a job candidate whether they’re gay, you’re not allowed to ask them whether they’re a college graduate or not. You can give them all sorts of examinations, you can ask them their high school grades and SAT scores, you can ask their work history, but if you ask them if they have a degree then that’s illegal class-based discrimination and you’re going to jail. I realize this is a blatant violation of my usual semi-libertarian principles, but at this point I don’t care.

# Michigan Meetup 6/13

There will be a Michigan rationalist/LW/SSC meetup at the Ann Arbor District Library (343 South Fifth Avenue, Ann Arbor) on Saturday, June 13 at 1:30 PM. You can find a map here.

If you’re reading this and can make it to Ann Arbor, you’re invited.

I will be in New York/New Jersey the weekend after that, and probably available the evening of Sunday the 21st. If anybody in the area has space for a meetup, let me know and we can work something out.

# (Late) Predictions for 2015

I was supposed to institute a tradition of making predictions at the beginning of each year, then grading them at the end to test my calibration. Everything went according to plan last year – last January I made predictions for 2014 and then this January I scored them. I ended with “2015 predictions coming soon,” then totally forgot to do that.

So now that it’s June and predicting what will happen in 2015 is about 40% easier, I might as well get started. Acceptable confidence levels are 50%, 60%, 70%, 80%, 90%, 95%, and 99%, to make it easy to score. All predictions are about the state of the world on 12/31/2015:

World Events  
1. US will not get involved in any new major war with death toll of > 100 US soldiers: 70%  
2. North Korea’s government will survive the year without large civil war/revolt: 95%  
3. Greece will not announce it’s leaving the Euro: 60%  
3. Neither Russia nor Qatar will lose their World Cups: 80%  
4. Ebola will kill fewer people in second half of 2015 than the in first half: 95%  
5. No terrorist attack in the USA will kill > 100 people: 90%  
6. Assad will remain President of Syria: 70%  
7. Israel will not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 90%  
8. Syria’s civil war will not end this year: 80%  
9. ISIS will control less territory than it does right now: 70%  
10. ISIS will continue to exist: 80%  
11. Iran will reach a deal with the West on nuclear weapons: 80%  
12. No major civil war in Middle Eastern country not currently experiencing a major civil war: 90%  
13. Iraq’s situation not to get any worse (eg gov’t collapse, new rebellion): 60%  
14. Obamacare will survive the year mostly intact: 60%  
15. Hillary Clinton will be the top-polling Democratic Presidential candidate: 95%  
16. Jeb Bush will be the top-polling Republican candidate: 50%  
17. Trans-Pacific Partnership to pass at least mostly intact: 60%  
18. US official unemployment rate will be less than 7% in Dec 2015: 95%  
19. Bitcoin will end the year higher than $200: 95%  
20. Oil will end the year greater than $60 a barrel: 50%

Personal Life  
21. SSC will remain active: 95%  
22. SSC will get fewer hits in the second half of 2015 than the first half: 60%  
23. At least one SSC post in the second half of 2015 will get > 100,000 hits: 70%  
24. Shireroth will remain active: 90%  
25. I will remain at my same job through the end of 2015: 95%  
26. There will be no further ramifications or lawsuits from either side over the flooding of my house: 80%  
27. I will reach my savings target: 90%  
28. I will get a score at >95th percentile for my year on PRITE: 50%  
29. I will be involved in at least one published/accepted-to-publish research paper by the end of 2015: 60%  
30. I will not break up with any of my current girlfriends: 80%  
31. I will not get any new girlfriends: 50%  
32. I will not finish [project]: 60%  
33. I will attend NYC Solstice ritual: 80%  
34. I will flake out of my plan to lead some kind of Solstice Ritual myself: 60%  
35. I will be living in the house I’m currently trying to arrange to rent: 70%

These are all the things I could think of worth predicting; feel free to suggest others.

# Fearful Symmetry

[Content warning: Social justice, anti-social justice, comparisons of social justice to anti-social-justice, comparisons of different groups’ experiences.]

The social justice narrative describes a political-economic elite dominated by white males persecuting anybody who doesn’t fit into their culture, like blacks, women, and gays. The anti-social-justice narrative describes an intellectual-cultural elite dominated by social justice activists persecuting anybody who doesn’t fit into their culture, like men, theists, and conservatives. Both are relatively plausible; Congress and millionaires are 80% – 90% white; journalists and the Ivy League are 80% – 90% leftist.

The narratives share a surprising number of other similarities. Both, for example, identify their enemy with the spirit of a discredited mid-twentieth century genocidal philosophy of government; fascists on the one side, communists on the other. Both believe they’re fighting a war for their very right to exist, despite the lack of any plausible path to reinstituting slavery or transitioning to a Stalinist dictatorship. Both operate through explosions of outrage at salient media examples of their out-group persecuting their in-group.

They have even converged on the same excuse for what their enemies call “politicizing” previously neutral territory – that what their enemies call “politicizing” is actually trying to restore balance to a field the other side has already successfully politicized. For example, on Vox recently a professor accused of replacing education with social justice propaganda in her classroom counterargues that:

All of my students, regardless of the identity categories they embraced, had been taught their entire lives that real literature is written by white people. Naturally, they felt they were being cheated by this strange professor’s “agenda”…It is worth asking, Who can most afford to teach in ways that are least likely to inspire controversy? Those who are not immediately hurt by dominant ideas. And what’s the most dominant idea of them all? That the white, male, heterosexual perspective is neutral, but all other perspectives are biased and must be treated with skepticism […]

Have we actually believed the lie that the only people who engage in “identity politics” are black feminists like me? Could it be that when some white men looked at more powerful white men, they could see them only as reasonable and not politically motivated, so they turned off their critical thinking skills when observing their actions? (Not everyone, of course.) Could it be that we only consider people ideologues when they don’t vow allegiance to capitalism?

Compare to the “Sad Puppies”, a group of conservatives accused of adding a conservative bent to science fiction’s Hugo Awards. They retort that “politicization is what leftists call it when you fight back against leftists politicizing something”. As per the Breitbart article:

The chief complaint from the Sad Puppies campaigners is the atmosphere of political intolerance and cliquishness that prevails in the sci-fi community. According to the libertarian sci-fi author Sarah A. Hoyt, whispering campaigns by insiders have been responsible for the de facto blacklisting of politically nonconformist writers across the sci-fi community. Authors who earn the ire of the dominant clique can expect to have a harder time getting published and be quietly passed over at award ceremonies […]

Brad R. Torgersen, who managed this year’s Sad Puppies campaign, spoke to Breitbart London about its success: “I am glad to be overturning the applecart. Numerous authors, editors, and markets have been routinely snubbed or ignored over the years because they were not popular inside WSFS or because their politics have made them radioactive.”

Torgersen cites a host of authors who have suffered de facto exclusion from the sci-fi community: David Drake, David Weber, L.E Modesitt Jr, Kevn J. Anderson, Eric Flint, and of course Orson Scott Card — the creator of the world-famous Ender’s Game, which was recently adapted into a successful movie. Despite his phenomenal success, Scott Card has been ostracized by sci-fi’s inner circle thanks to his opposition to gay marriage.

I see minimal awareness from the social justice movement and the anti-social-justice movement that their narratives are similar, and certainly no deliberate intent to copy from one another. That makes me think of this as a case of convergent evolution.

The social justice attitude evolved among minority groups living under the domination of a different culture, which at best wanted to ignore them and at worst actively loathed them for who they were and tried to bully them into submission. The closest the average white guy gets to that kind of environment is wandering into a social-justice-dominated space and getting to experience the same casual hatred and denigration for them and everyone like them, followed by the same insistence that they’re imagining things and how dare they make that accusation and actually everything is peachy.

And maybe that very specific situation breeds a very specific kind of malignant hypervigilance, sort of halfway between post-traumatic stress disorder and outright paranoia, which motivates the obvious fear and hatred felt by both groups.

Someone is going to freak out and say I am a disgusting privileged shitlord for daring to compare the experience of people concerned about social justice to the experience of genuinely oppressed people, but they really shouldn’t. That’s the explicit goal of large parts of the social justice movement. For example, on the Hacker News thread about far-rightist Curtis Yarvin being kicked out of a tech conference for his views, one commenter writes:

I’ve been involved in anti-racist/anti-fascist work, either directly or on the periphery, for about ten years at this point. This takes many forms, from street confrontations with fascists, protests at book readings and other events, and also disrupting fascist conferences and similar […]

As far as this issue and other similar issues are concerned, I’m overjoyed that, as you put it, a climate of fear exists for fascists, misogynists, racists, and similar. I hope that this continues and only worsens for these people.

I’m happy for many reasons. The first is that it has, as you’ve said, made privileged people afraid. I think this is only the beginning. Privilege creates safety, and as it is removed, I think the unsafety of the oppressed will in part come to the currently privileged classes. But if I could flip a switch and make every man feel the persistent, gnawing fear that a woman has of men, I would in a heartbeat. I wouldn’t even consider whether the consequences were strategic, I would just do it.

This not the only time I’ve heard this opinion expressed, just the most recent. I feel like if you admit that you’re trying your hardest to make privileged people feel afraid and uncomfortable and under siege in a way much like minorities traditionally do, and privileged people are in fact complaining of feeling afraid and uncomfortable and under siege in a way much like minorities traditionally do, you shouldn’t immediately doubt their experience. Give yourself some more credit than that. You’ve been working hard, and at least in a few isolated cases here and there it’s paid off.

The commenter continues:

I would not say that I set out to defeat a “discourse-stifling” monster. The monsters I set out to defeat were patriarchy, capitalism, and white supremacy. These systems violently oppress, they don’t “stifle discourse.” In fact, they LOVE discourse! When people are discoursing, they aren’t in the streets. I’ve seen so many promising movements hobbled by reformism that I’m glad the possibility no longer exists, though that isn’t at all the fault of SJW-outrage (and is rather a consequence of the fact that the economy is in large part so perilous that nobody can afford the concessions that were previously won by reformists). So if discourse is permanently removed as a tactical and strategic option for future leftists, I’ll consider it a victory.

Needless to say, that is not this blog’s philosophy. But I think there is nevertheless something to be gained from all of the hard work this guy and his colleagues have put in making other people feel unsafe.

The mirror neuron has always been one of liberalism’s strongest weapon. A Christian doesn’t decide to tolerate Islam because she likes Islam, she decides to tolerate Islam because she can put herself in a Muslim’s shoes and realize that banning Islam would make him deeply upset in the same way that banning Christianity would make her deeply upset.

If the fear and hypervigilance that majority groups feel in social-justice-dominated spaces is the same as the fear and hypervigilance that minority groups feel in potentially discriminatory spaces, that gives us a whole lot more mirror neurons to work with and allows us to get a gut-level understanding of the other side of the dynamic. It lets us check my intuitions against their own evil twins on the other side to determine when we are proving too much.

II.

A couple of months ago the owners of a pizzeria mentioned in an interview that they wouldn’t serve pizza at gay weddings because they’re against gay marriage. Instantly the nation united in hatred of them and sent a bunch of death threats and rape threats and eventually they had to close down.

I thought this was ridiculous. I mean, obviously death threats are never acceptable, but there seemed to be something especially frivolous about this case, where there are dozens of other pizzerias gay people can go to and where no one would ever serve pizza at a wedding anyway. A pizzeria hardly holds the World Levers Of Power, so just let them have their weird opinion. All they’re doing is sending potential paying customers to their more tolerant competitors, who are laughing all the way to the bank. It’s a self-punishing offense.

This was very reasonable of me and I should be praised for my reasonableness, except that when a technology conference recently booted a speaker for having far-right views on his own time, I was one of the many people who found this really scary and thought they needed to be publicly condemned for this intolerant act.

In theory, the same considerations ought to apply. There are dozens of other technology conferences in the world. Technology conferences also do not hold the World Levers Of Power. And when they reject qualified rightist speakers, that just means they’re just making life easier for their competitors who will be happy to grab the opportunity and laugh all the way to the bank. It ought to be self-punishing, so what’s the worry.

My brain is totally not on board with this reasoning. When I ask it why, it says something like “No, you don’t understand, these people are relentless, unless they are constantly pushed against they will put pressure on more and more institutions until their enemies are starved out or limited to tiny ghettos. Then they will gradually expand the definition of ‘enemy’ until everybody who doesn’t do whatever they say is blacklisted from everywhere.”

And if you think that’s hyper-paranoid, then, well, you’re probably right, but at least I have a lot of company. Here are some other comments on the same situation from the last links thread:

I spent a semester of college in Massachusetts. That’s where I found out that there are a lot of people who’d kill me and most of my family if they were given the chance. And thought it was totally reasonable and acceptable to say as much. (The things that are associated with Tumblr these days existed long before it. And mostly came from academia.)

About the same time that sort of thing was happening in that online community, the same thing was happening in the real-world meat-space gatherings, also quite literally with shrill screams, mostly by [reacted] [reacted]s, who would overhear someone else’s private conversations, and then start streaming “I BEG YOUR PARDON!” and “HOW CAN YOU SAY THAT!”, and by [reacted] [reacted]’s who were bullying their way onto programming committees, and then making sure that various speakers, panelists, artists, authors, dealers, and GoHs known to be guilty of wrongthink were never invited in the first place. Were it not for the lucky circumstance of the rise of the web, the market takeoff of ebooks, especially a large ebook vendor (named after a river)’s ebook direct program, and the brave anchoring of a well known genre publisher that was specifically not homed in NYC, the purging of the genre and the community would have been complete.

Almost nobody wants to physically murder and maim the enemy, at least at the start. That’s, well, the Final Solution. Plan A is pretty much always for the enemy to admit their wrongness or at least weakness, surrender, and agree to live according to the conqueror’s rules. Maybe the leaders will have to go to prison for a while, but everyone else can just quietly recant and submit, nobody has to be maimed or killed. [The social justice community] almost certainly imagine they can achieve this through organized ostracism, social harassment, and democratic political activism. It’s when they find that this won’t actually make all the racists shut up and go away, that we get to see what their Plan B, and ultimately their final solution, look like.

And if you think my commenters are also hyper-paranoid, then you’re probably still right. But it seems like the same kind of paranoia that makes gay people and their allies scream bloody murder against a single pizzeria, the kind that makes them think of it as a potential existential threat even though they’ve won victory after victory after victory and the only question still in the Overton Window is the terms of their enemies’ surrender.

I mocked the hell out of the people boycotting Indiana businesses because of their right-to-discriminate law:

Can we admit it's KIND OF funny ppl are boycotting Indiana for the immoral act of allowing people to boycott those they think act immorally?

— Scott Alexander (@slatestarcodex) March 31, 2015

But if some state were to pass a law specifically saying “It is definitely super legal to discriminate against conservatives for their political beliefs,” this would freak me out, even though I am not conservative and even though this is already totally legal so the law would change nothing. I would not want to rule out any response, up to and including salting their fields to make sure no bad ideas could ever grow there again.

Like many people, I am not very good at consistency.

III.

Author John Green writes books related to social justice. A couple of days ago, some social justice bloggers who disagreed with his perspective decided that a proportional response was to imply he was a creep who might sexually abuse children. Green was somewhat put out by this, and said on his Tumblr that he was “tired of seeing the language of social justice – important language doing important work – misused as a way to dehumanize others and treat them hatefully” and that he thought his harassers “were not treating him like a person”.

Speaking of the language of social justice, “dehumanizing” and “not treating like a person” are some pretty strong terms. They’re terms I’ve criticized before – like when feminists say they feel like women aren’t being treated as people, I’m tempted to say something like “the worst you’ve ever been able to find is a single-digit pay gap which may or may not exist, and you’re going to turn that into people not thinking you’re human?”

Here’s another strong term: “hatred”. The activist who got Mencius Moldbug banned from Strange Loop reassured us that he would never want someone banned merely for having unusual political views, but Moldbug went beyond that into “hatred”, which means his speech is “hate speech”, which is of course intolerable. This is a bit strange to anybody who’s read any of his essays, which seem to have trouble with any emotion beyond smugness. I call him a bloodless and analytical thinker; the idea of his veins suddenly bulging out when he thinks about black people is too silly to even talk about. The same is true of the idea that people should feel “unsafe” around him; his entire shtick is that no one except the state should be able to initiate violence!

Likewise, when people wanted TV star Phil Robertson fired for saying (on his own time) that homosexuality was unnatural and led to bestiality and adultery, they said it wasn’t about policing his religion, it was about how these were “hateful” comments that would make the people working with him feel unsafe. At the time I said that was poppycock and that people who wanted him fired for having a private opinion were the worst kinds of illiberal witch-hunters.

On the other hand, consider Irene Gallo. I know nothing of her except what the Alas blog post says, but apparently in science fiction’s ongoing conflict between the establishment and the anti-SJW “Sad Puppies”/”Rabid Puppies” groups, she referred to the latter as:

Two extreme right-wing to neo-nazi groups that are calling for the end of social justice in science fiction and fantasy. They are unrepentantly racist, sexist and homophobic.

These are some pretty strong allegations, and range from “false” to “bizarre”; Brad Torgenson, leader of the group she called “extreme right wing neo nazi unrepentant racists”, is happily married to a black woman. And the people she’s talking about are her company’s authors and customers, which hardly seems like good business practice. Some authors have said they feel uncomfortable working for a company whose employees think of them that way, and others have suggested boycotting Tor until they make her apologize or fire her.

Barry says that since she said these on her own private Facebook page, it is a private opinion that it would be pretty censorious to fire her over. Part of me agrees.

On the other hand, if I were a sci-fi author in one of the groups that she was talking about, I’m not sure I’d be able to work with her. Like, really? You want me to sit across a table and smile at the woman who thinks I’m a racist sexist homophobic extremist neo-Nazi just because I disagree with her?

Robertson’s comment is just standard having-theological-opinions. Like, “Christian thinks homosexuality is sinful, more at eleven.” Big deal. But Gallo’s comment feels more like white hot burning hatred. She’s clearly too genteel to personally kill me, but one gets the clear impression that if she could just press a button and have me die screaming, she’d do it with a smile on her face.

But this is just interpretation. Maybe Gallo doesn’t consider “neo-Nazi” a term of abuse. Maybe this was just her dispassionate way of describing a political philosophy with the most appropriate analogy she could think of.

It doesn’t seem likely to me. Then again, even though it seems obvious to me that stating “homosexuality is sinful and similar to bestiality” is a theological position totally compatible with being able to love the sinner and hate the sin, gay people have a lot of trouble believing it. And although I cannot condone firing people for their private opinions, back when people were trying to get rid of Gawker honcho Sam Biddle for saying that “nerds should be constantly shamed and degraded into submission”, God help me it certainly crossed my head that there were even the slightest consequences for this kind of behavior, maybe other social justice writers would stop saying and acting upon statements like that all the frickin’ time?

Once again, I’m not scoring very highly in consistency here.

IV.

A little while ago I had a bad couple of days. Some people were suggesting I was a liability to a group I was part of because I’d written some posts critical of feminism, and I got in a big fight about it. Then someone sent my ex-girlfriend a Tumblr message asking if they’d broken up with me “because I was racist”. Then despite my best efforts to prevent this, my Facebook feed decided to show me a bunch of Gawker-style articles about “Are all white people to blame for [latest atrocity]? I was too exhausted to write a real blog post, so I just threw together a links post. Because among two dozen or so links there was one (1) to the Moldbug story previously mentioned above, one commenter wrote that “your links posts are becoming indistinguishable from Chaos Patch” (Chaos Patch is the links post of notable far-right blog Xenosystems).

So I decided to ban that commenter. But since I have a policy in place of waiting an hour before doing anything rash, I took a long walk, thought about it a bit, and settled for just yelling at him instead.

Is banning someone for a kind of meaningless barb excessive? Well, yes. But given everything else that had happened, I didn’t have the energy to deal with it, and since this is my blog and the one corner of the world I have at least a tiny bit of control over I could at least symbolically get rid of a small fraction of my problems.

Plus, to me the barb seemed like an obvious veiled threat. “As long as you post any links about rightist causes, I can accuse you of being far-right. And we all know what happens to far-right people, eh?”

So even though out of context it was about the most minimal hostility possible, barely rising to the level where somebody would say it was even capable of being a problem at all, in context it really bothered me and made me at least somewhat justifiably feel unsafe.

Ever since I learned the word “microaggression” I have been unironically fond of it.

Microagressions. Nanoagressions. Picoagressions. The Planck Hostility.

— Map of Territory (@MapOfTerritory) January 28, 2015

When I’m putting up with too much and I’ve used up my entire mental buffer, then somebody bothering me and hiding under the cover of “oh, this was such a tiny insult that you would seem completely crazy to call me on it” is especially infuriating, even more infuriating than someone insulting me outright and me being able to respond freely. The more you have to deal with people who hate you and want to exclude you, the more likely you are to get into this mode, not to mention people who have developed their own little secret language of insults.

Here’s an example of what I mean by “secret language of insults”: consider the term “dude”, as in “white dude”. There is nothing objectively wrong with “dude” when it is applied to surfers or something. But when a feminist says it, as in the term “white dudes”, you know it is going to be followed by some claim that as a white dude, you are exactly the same as all other white dudes and entirely to blame for something you don’t endorse. The first page of Google results is overratedwhitedudes.tumblr.com, Gawker saying Wimpy White Dudes Ruined American Idol, and Mother Jones saying glowingly that You Won’t Find Many White Dudes At This Tech Startup. Being called a “white dude” is always followed by the implication that you’re ruining something or that your very presence is cringeworthy and disgusting.

I had a feminist friend who used to use the term “dudes” for “men” all the time. I asked them to please stop. They said that was silly, because that was just the word the culture they’d grown up in used, and obviously no harm was meant by it, and if I took it as an insult then I was just being oversensitive. This is word for word the explanation I got when I asked one of my elderly patients to stop calling black people their particular ethnic slur.

The counterpart to subliminal insults is superliminal insults; ones that are hard to detect because they’re so over-the-top obvious.

I was recently reading a social justice blog where someone complained about men telling women “Make me a sandwich!” in what was obvious jest.

On the one hand, no one can possibly take this seriously.

On the other hand, there’s a common social justice meme where people post under the hashtag #killallwhitemen.

Certainly this cannot be taken seriously; most social justice activists don’t have the means to kill all white men, and probably there are several of them who wouldn’t do it even if they could. It should not be taken, literally, as a suggestion that all white men should be killed. On the other hand, for some bizarre reason this tends to make white men uncomfortable.

The obvious answer is that the people posting “Wimmen, make me a sandwich!” don’t literally believe that women exist only for making them sandwiches, but they might believe a much weaker claim along the same lines, and by making the absurd sandwich claim, they can rub it in while also claiming to be joking. At least this is how I feel about the “kill all white men” claim.

As long as you’ve got a secret language of insults that your target knows perfectly well are insulting, but which you can credibly claim are not insulting at all – maybe even believing it yourself – then you have the ability to make them feel vaguely uncomfortable and disliked everywhere you go without even trying. If they bring it up, you can just laugh about how silly it is that people believe in “microaggressions” and make some bon mot about “the Planck hostility”.

V.

I’m taking a pretty heavy Outside View line here, so let me allow my lizard brain a few words in its own defense.

“Yes,” my lizard brain says, “social justice activists and the people silenced by social justice activists use some of the same terms and have some of the same worries. But the latter group has reasonable worries, and the former group has totally unreasonable worries, which breaks the symmetry.”

Interesting. Please continue, lizard brain.

“Black people might be very worried about being discriminated against. But the chance that someone would say ‘Let’s ban all black people from our technology conference, because they are gross’, and everyone would say ‘Yes, that is a splendid idea’, and the government and media would say ‘Oh, wonderful, we are so proud of you for banning all black people from your conference’ is zero point zero zero zero. On the other hand, this is something that conservatives worry about every day. The chance that someone would say ‘You know, there’s no reason raping women should be illegal, let’s not even bother recording it in our official statistics’ is even lower than that, but this is exactly what several countries do with male rape victims. If someone says ‘kill all white men’, then all we do is hold an interminable debate about whether that disqualifies them from the position of Diversity Officer; if someone said ‘kill all gays’, we would be much more final in pronouncing them Not Quite Diversity Officer Material.”

But don’t you –

“The reason why we don’t care about a pizzeria that won’t serve gay people is that recent years have shown an overwhelming trend in favor of more and more rights and acceptance of gay people, and the pizzeria is a tiny deviation from the pattern which is obviously going to get crushed under the weight of history even without our help. The reason we worry about a conference banning conservatives is that conservatives are an actually-at-risk group, and their exclusion could grow and grow until it reaches horrific proportions. The idea of a pizzeria banning gays and a conference banning conservatives may seem superficially similar out of context, but when you add this piece of context they’re two completely different beasts.”

Two responses come to mind.

First, this is obviously true and correct.

Second, this is exactly symmetrical to my least favorite argument, the argument from privilege.

The argument from privilege is something like “Yeah, sure, every so often the system is unfair to white people or men or whatever in some way. But this is not a problem and we should not even be talking about it, because privilege. Shows that mock women for stereotypically female failings are sexist, but shows that mock men for stereotypically male failings are hilarious, and you may not call them sexist because you can’t be sexist against privileged groups.”

My argument has always been “What’s good for the goose is good for the gander”.

But either this argument goes, or my lizard brain’s argument goes, or we have to move to the object level, or somebody has to get more subtle.

VI.

My point is, there are a lot of social justice arguments I really hate, but which I find myself unintentionally reinventing any time things go really bad for me, or I feel like myself or my friends are being persecuted.

I should stop to clarify something. “Persecuted” is a strong word. “Feel like we are being persecuted” is way weaker.

A couple weeks ago there was a Vox article, America’s Never Been Safer, So Why Do Republicans Believe It Is In Mortal Peril?. It brought up a lot of cute statistics, like that the rate of pedestrians being killed by car accidents is much higher than the rate of civilians being killed in terrorist attacks. It joked that “You’re over 100 times more likely to die by literally walking around than you are to be killed in a terrorist attack.”

On the other hand, vox has practically led the news media in 24-7 coverage of police officers shooting unarmed black people, talking about how it’s a huge threat to our values as a civilization and how white people don’t understand that all black people have to constantly live in fear for their lives.

But a quick calculation demonstrates that unarmed black people are about 10 times more likely to die by literally walking around than by getting shot by a white police officer. One gets the feeling Vox doesn’t find this one nearly as funny.

But here I would perform another quick calculation. Here’s a list of people who have been publicly shamed or fired for having politically incorrect opinions. Even if we assume the list is understating the extent of the problem by an entire order of magnitude, you’re still more likely to die by literally walking around than you are to get purged for your politically incorrect opinion.

Like a lot non-feminists, I was freaked out by the recent story about a man who was raped while unconscious being declared the rapist and expelled from college without getting to tell his side of the story. I have no evidence that this has ever happened more than just the one time mentioned in the article, let alone it being a national epidemic that might one day catch me in its clutches, but because I’ve had to deal with overly feminist colleges in other ways, my brain immediately raised it to Threat Level Red and I had to resist the urge to tell my friends in colleges to get out while they still could. If we non-feminists can get worried about this – and we can – we have less than no right to tell feminists they shouldn’t really be worried about college rape because the real statistics are 1 in X and not 1 in Y like they claim.

Hopefully some readers are lucky enough never to have felt much personal concern about terrorism, police shootings, rape, rape accusations, or political correctness. But if you’ve worried about at least one of these low-probability things, then I hope you can extend that concern to understand why other people might be worried about the others. It seems to have something to do with the chilling effect of knowing that something is intended to send a message to you, and in fact receiving that message.

(as an aside, I find it surprising that so many people, including myself, are able to accept the statistics about terrorism so calmly without feeling personally threatened. My guess is that, as per Part VIII here, we don’t primarily identify as Americans, so a threat deliberately framed as wanting to make Americans feel unsafe just bounces off us.)

In an age where the media faithfully relates and signal-boosts all threats aimed at different groups, and commentators then serve their own political needs by shouting at us that WE ARE NOT FEELING THREATENED ENOUGH and WE NEED TO FEEL MORE THREATENED, it is very easy for a group that faces even a small amount of concerted opposition, even when most of society is their nominal allies and trying hard to protect them, to get pushed into a total paranoia that a vast conspiracy is after them and they will never be safe. This is obviously the state that my commenters who I quoted in Part II are stuck in, obviously the state that those people boycotting the Indiana pizzeria are stuck in, and, I admit, a state I’m stuck in a lot of the time as well.

VII.

Getting back to the thesis, my point is there are a lot of social justice arguments I really hate, but which I find myself unintentionally reinventing any time things go really bad for me, or I feel like myself or my friends are being persecuted.

Once events provoke a certain level of hypervigilance in someone – which is very easy and requires only a couple of people being hostile, plus the implication that they there’s much more hostility hidden under the surface – then that person gets in fear for their life and livelihood and starts saying apparently bizarre things: that nobody treats them as a person, that their very right to exist is being challenged. Their increasingly strident rhetoric attracts increasingly strident and personal counter-rhetoric from the other side, making them more and more threatened until they reach the point where Israel is stealing their shoe. And because they feel like every short-term battle is the last step on the slippery slope to their total marginalization, they engage in crisis-mode short-term thinking and are understandably willing to throw longer-term values like free speech, politeness, nonviolence, et cetera, under the bus.

Although it’s very easy enter this state of hypervigilance yourself no matter how safe you are, it’s very hard to understand why anyone else could possibly be pushed into it despite by-the-numbers safety. As a result, we constantly end up with two sides both shouting “You’re making me live in fear, and also you’re making the obviously false claim that you live in fear yourself! Stop it!” and no one getting anywhere. At worst, it degenerates into people saying “These people are falsely accusing me of persecuting them, and falsely claiming to be persecuted themselves, I’ll get back at them by mocking them relentlessly, doxxing them, and trying to make them miserable!” and then you get the kind of atmosphere you find in places like SRS and Gamergate and FreeThoughtBlogs.

But I’m also slightly optimistic for the future. The conservative side seems to have been about ten years behind the progressive side in this, but they’re catching up quickly. Now everybody has to worry about being triggered, everybody has to worry about their comments being taken out of context by Gawker/Breitbart and used to get them fired and discredit their entire identity group, everybody has to worry about getting death threats, et cetera. This is bad, but also sort of good. When one side has nukes, they nuke Hiroshima and win handily. When both sides have nukes, then under the threat of mutually assured destruction they eventually come up with protocols to prevent those nukes from being used.

Now that it’s easier to offend straight white men, hopefully they’ll agree trigger warnings can be a useful concept. And now that some social justice activists are getting fired for voicing their opinions in private, hopefully they’ll agree that you shouldn’t fire people for things they say on their own time. Once everyone agrees with each other, there’s a chance of getting somewhere. Yes, all of this will run up against a wall of “how dare you compare what I’m doing to what you’re doing, I’m defending my right to exist but you’re engaging in hate speech!” but maybe as everyone gets tired of the nukes flying all the time people will become less invested in this point and willing to go to the hypothetical Platonic negotiation table.

My advice for people on the anti-social justice side – I don’t expect giving the SJ people advice would go very well – is that it’s time to stop talking about how social justice activism is necessarily a plot to get more political power, or steal resources, or silence dissenting views. Like everything else in the world it can certainly turn into that, but I think our own experience gives us a lot of reasons to believe they’re exactly as terrified as they say, and that we can’t expect them to accept “you have no provable objective right to be terrified” any more than our lizard brains would accept it of us. I think it’s time to stop believing that they censor and doxx and fire their opponents out of some innate inability to understand liberalism, and admit that they probably censor and doxx and fire their opponents because they’re as scared as we are and feel a need to strike back.

This isn’t a claim that they don’t have it in for us – many of them freely admit they do – and that they don’t need to be stopped. It’s just a claim that we can gain a good understanding of why they have it in for us, and how we might engineer stopping them in a way less confrontational than fighting an endless feud.

Yesterday, a friend on Facebook posted something about a thing men do which makes women feel uncomfortable and which she wanted men to stop. I carefully thought about whether I ever did it, couldn’t think of a time I had, but decided to make sure I didn’t do it in the future.

I realized that if I’d heard the exact same statement from Gawker, I would have interpreted it (correctly) as yet another way to paint men as constant oppressors and women as constant victims in order to discredit men’s opinions on everything, and blocked the person who mentioned it to me so I didn’t have to deal with yet another person shouting that message at me. The difference this time was that it came from an acquaintance who was no friend of feminism, who has some opinions of her own that might get her banned from tech conferences, and who I know would have been equally willing to share something women do that bothers men, if she had thought it important.

If we can get to a point where we don’t feel like requests are part of a giant conspiracy to discredit and silence us, people are sometimes willing to listen. Even me.

# Fifty (More) Swifties

[see: Wikipedia: Tom Swifties, Tom Swifties Written By An Author Willing To Go To Any Lengths To Make A Tom Swifty Thus Resulting In Constructions That Often Require More Work For Readers Than For The Author, and Fifty Swifties. Previously on Twitter here.]

“This sandwich is gross,” Tom said deliberately.

“My Frisbee is stuck on the roof of that circus building,” Tom said discontently.

“I hate Google,” Tom said probingly.

“Godzilla swallowed a United Nations bunker, but then he threw it back up,” Tom said unfortunately.

“I think Objectivism is stupid,” Tom said randomly.

“It’s so exciting to visit Leonardo’s birthplace,” Tom said invincibly.

“Persephone must marry Hades and live with him half the year,” Zeus said despairingly.

“I now control majority shares of CBS, FOX, and the New York Times,” Tom said immediately.

“Enemy fighters just scored a direct hit on my plane! I’m going down!” Tom said knowingly.

“We were badly injured in the struggle with the Orcs, but luckily the Ents’ medicine restored our health,” Tom said tremendously.

“I took Gollum’s precious trinket in a riddle contest,” Tom said wonderingly.

“I’ve lost this Maxis game ten times in a row on the easiest difficulty setting,” Tom said sympathetically.

“I can commit adultery three more times and still be just under the threshold for damnation,” Tom said syntactically.

“O Lord, why are you punishing me like this?” Jonah said inefficiently.

“Look! Nicaraguan guerillas!” Tom contraindicated.

“I forgot to give up meat before Easter, so I’ll do it before Christmas,” Tom said redolently.

“I’ll see you in court!” Tom said supersonically.

“When I speak Japanese, I think of myself as a young, cute person,” Tom said mechanically.

“Iä Cthulhu! Iä Azathoth!” the man called maniacally.

“Stay away from Stalin,” Tom commissioned.

“It’s one of those old phones, from before wireless and touch-tone,” Tom said cordially.

“She’ll have sex with me for $20 any time I phone her up,” Tom said horrifically.

“I read the Cliff Notes to Dante’s Inferno,” Tom said synergistically.

“I’m going to recover the lunar lander from the surface of the moon and make a fortune,” Tom said apologetically.

“I covered myself in a layer of gold,” Tom said amblingly.

“I covered myself in a layer of pyrite,” Tom said shamblingly.

“I covered myself in the Golden Fleece of Colchis,” Tom said ramblingly.

“The poverty rate has increased 10% recently, but I don’t have any kind of visual presentation of its course,” Tom said pornographically.

“We should perform an autopsy,” Tom said wide-eyed.

“That tree is naked under its bark!” Tom said prudently.

“I can afford either an iPhone or a yacht, but not both,” Tom said on self-ownership.

“The guy who was installing the granite tops in my kitchen had a cardiac arrest,” Tom countermanded.

“We can stop progress by attacking a conference on new ideas with a many-headed monster,” Tom said well-hydrated.

“You’re a bell,” Tom told me.

“The wages of sin is death,” Tom said diurnally.

“Abortion is murder,” Tom said prolifically.

“Can do!” Tom said candidly.

“I have a present for you, Madame,” Vincent said endearingly.

“Arrrrrr,” Tom aspirated.

“My lower social status as part of the new rich prevents me from winning my true love,” Gatsby said lackadaisically.

“The Minoans sucked,” Tom said discretely.

“Well, if you think the Minoans did a bad job with their empire, you should try ruling them yourself,” his teacher said, giving him a B−.

“Ha ha, just kidding,” Tom ingested.

“Sheep can’t have sex changes!” Tom said, heedless of the ramifications.

“I wrote a synoptic Gospel,” Tom remarked.

“People used to lay wires across the country for the telegraph system, an early precursor to the telephone,” Tom said according to protocol.

“My laptop came bundled with malware that causes a serious security flaw,” Tom said superficially.

“We need artillery cover!” Tom said canonically.

“Someday my family will rule the world,” Tom said clandestinely.

“The West’s treatment of Palestine is an example of Orientalism,” Tom said.

# NYC Meetup 6/21

There will be a New York City rationalist/LW/SSC meetup at Highgarden House (851 Park Place, Brooklyn) on Sunday, June 21 at 6:00 PM.

If you’re reading this and can make it to Brooklyn, you’re invited.

There was some discussion at the last Michigan meetup about how the previous sentence should be emphasized more. If you are reading this and can make it to Brookyln, you’re invited. You do not have to have read very much of this blog or Less Wrong, you do not have to agree with me about anything, you do not have to “be smart enough” for any value of “enough”, and you do not have to be the “sort of person” whom you think goes to these things. In fact, the less you are like everyone else, the more interesting it will be.

No particular activities planned beyond undirected conversation, but there will probably be pizza, and Raemon has raised the possibility of a dramatic reading of A Midsummer Night’s Dream for those interested (in a way that doesn’t interfere with continued undirected conversation for people who don’t want to do that).

# Reflections From The Halfway Point

I.

A while back one of my patients was having a foot problem, so I consulted the hospital podiatrist. He met me in my workroom, and I explained exactly what I needed from him, but over the course of the explanation he started looking more and more uncomfortable and distracted, so finally I stopped and was just like “Okay, out with it, what’s your problem?”

And he said: “That guy with the wild hair pounding on the window and shouting threats and obscenities at us.”

And I said: “Oh, him? That’s just Bob. Don’t worry about him, he always does that.”

The podiatrist seemed inadequately reassured.

I thought about this because as of today I am halfway done with my four-year psychiatry residency.

One of my teachers told me that you go to medical school to learn things, and then you go to residency to get used to them. It’s not quite that simple – you certainly learn a lot in residency – but there’s a lot of truth to it. I remember that my first week on call, somebody had a seizure and I totally freaked out – AAAAH SEIZURE WHAT DO I DO WHAT DO I DO? – even though I had previously been able to pass tests on that exact situation. But my last time on call, somebody also had a seizure, and I sort of strolled in half-asleep, ordered the necessary tests and consultations and supportive care, then strolled out and went back to bed.

And then there are the little things, like learning to tune out a psychotic guy banging on the window and yelling threats at you.

II.

It’s interesting that psychiatric hospitals are used as a cliche for “a situation of total chaos” – I think I’ve already mentioned the time when the director of a psych hospital I worked at told us, apparently without conscious awareness or irony, that if Obamacare passed our hospital would have too many patients and “the place would turn into a madhouse”. There’s a similar idiom around “Bedlam”, which comes from London’s old Bethlehem psychiatric hospital.

In fact, psych hospitals are much more orderly than you would think. Maybe 80% of the patients are pretty ‘with it’ – depressed people, very anxious people, people with anger issues who aren’t angry at the moment, people coming off of heroin or something. The remaining 20% of people who are very psychotic mostly just stay in their rooms or pace back and forth talking to themselves and not bothering anyone else. The only people you really have to worry about most of the time are the manic ones and occasionally severe autistics, and even they’re usually okay.

For a place where two dozen not-very-stable people are locked up in a small area against their will, violence is impressively rare. The nurses have to deal with some of it, since they’re the front-line people who have to forcibly inject patients with medication, and they have gotten burned a couple of times. And we doctors are certainly trained to assess for it, defuse it, and if worst comes to worst hold our own until someone can get help.

Yet in the two years I’ve worked at Our Lady Of An Undisclosed Location, years when each doctor has talked to each of their patients at least once a day, usually alone in an office, usually telling them things they really don’t want to hear like “No, you can’t go home today” – during all that time, not one doctor has been attacked. Not so much as a slap or a poke.

I am constantly impressed with how deeply the civilizing instinct has penetrated. When I go out of the workroom and tell Bob, “I’m sorry, but you’re disturbing people, you’re going to have to stop banging on the window and shouting threats, let’s go back to your room,” then as long as I use a calm, quiet, and authoritative voice, that is what he does. With very few exceptions, there is nobody so mentally ill that calmness + authority + the implied threat of burly security guards won’t get them to grumble under their breath but generally comply with your requests, reasonable or otherwise.

III.

I’d like to say I’ve taken advantage of this to go mad with power. But it’s actually a really crappy situation for everyone involved.

The most common reason for admission to a psychiatric hospital is “person is a danger to themselves or others”. The average length of stay in a psychiatric hospital is about one week.

Some clever person might ask: “Hey, don’t most psychiatric medicines require more than a week to take effect?” Good question! The answer is “yes”. Antidepressants classically take four weeks. Lithium and antipsychotics are more complicated, but the textbooks will still tell you a couple of weeks in both cases. And yet people are constantly being brought to psychiatric hospitals for dangerousness, treated with medications for one week, and then sent off. What gives?

As far as I can tell, a lot of it is the medical equivalent of security theater.

The most common type of case I see is “person who was really angry, said ‘I’ll kill myself’ in a fit of rage, and then their partner called the cops and they were brought to hospital.” These people stop being angry after a day or two and then no longer make these comments, even assuming they meant it in the first place which most of them don’t.

The second most common type of case I see is “person who was really angry, did try to kill themselves, and it didn’t work.” Again, these people have stopped being angry. Failed suicide attempts also have their own interesting way of clearing the mind for a little while, so they’re in a sort of grace period. Sending these people to a psychiatric hospital makes the public feel good because they’re Doing Something About Suicide, and makes psychiatrists feel good because after a few days they’ve stopped being suicidal so it looks like we’re Making A Difference. There is no way we could leave this equilibrium now even if we wanted to, because if we didn’t keep these people for a week and they ever attempted suicide again, we would get sued to oblivion.

The third most common type of case I see is “severely mentally ill person who’s been living at a care home for twenty years, but then they got in a fight and so their care home sent them to the hospital.” We shuffle their medications around and send them back to the care home where they’d been living happily for twenty years until some random trigger set them off.

We don’t call this “security theater”. We do sometimes call it a “holding environment”. Psych hospitals are kind of boring. There’s no boyfriend to get in a screaming match with, no boss pushing you to work harder, and no drug dealers to get heroin from. On the other hand, there’s lots of structure – art therapy at 10, meeting with your doctor at 11, recreation group at 12, and so on. It’s like a terrible vacation in the world’s least attractive hotel. People get a chance to cool off and forget about whatever set them off. Then they go back to their life. If they’re lucky, our social workers have managed to connect them to a better outpatient psychiatrist, care home, or support group, and maybe that will improve their lives sometime down the line. But I don’t think anyone imagines there was some fundamental Quality Of Dangerousness in them which is now gone.

To the degree that it is all security theater, it’s really hard to give an honest answer to a patient asking why they have to stay in hospital.

When I first started this work, my reaction to these people was “Come on, it’s only a week, it’s not like you’re stuck here forever, just deal with it.” This lasted until I remembered that when some stupid policy forces me to come into hospital on a day I would otherwise have off, I freak out, because I value my free time too much to be okay with having it taken away from me for bad reasons. Heck, my power was out the past couple of days, and I couldn’t use the Internet, and I was calling the power company and being like “COME ON YOU NEED TO FIX THIS ALREADY I AM LOSING DAYS OF MY LIFE THAT I COULD OTHERWISE BE SPENDING IN IMPORTANT STUFF.” So now I try to avoid throwing stones.

(there’s another aspect of this, which is that people constantly protest that horrible things will happen to them based on that week. For example: “My boss said if I miss one more day of work, I’ll lose my job, and then I’ll have no way to support my family.” Or: “My rent payment is due tomorrow, if I miss it I’ll be evicted and all of my stuff will go to the landfill, and there’s no way I can handle this through Internet or telephone or asking a friend to help.” I assume 90% of these stories are false, but the 10% that are true are still bad enough to more than outbalance any good we can do.)

After that, my reaction to these people was “Yes, you may be angry now, but you will thank us later.” This is true of many people, including some of the most histrionically upset. But I’ve since learned that it’s probably not true of the majority. The Shrink Rap blog surveyed former psychiatric inpatients and found that 62% said their experience was not helpful and they were “the same or worse at discharge”. I’d like to dismiss this as people just carrying a grudge for having to be there at all, but the same survey finds that a very similar 56% of voluntarily admitted patients said the same thing (although not all “voluntary” admissions are as voluntary as the name expects). Now, I don’t know for sure what to think about that survey – a lot of people describe their hospitals as doing things which are super illegal and which I wouldn’t expect a hospital to be able to get away with and stay open for more than twenty-four hours, and the population of psych patients who read psychiatric blogs is probably a nonrandom sample – but I no longer feel like I can confidently say that our patients will thank us later.

(none of this is to say that you shouldn’t check yourself into a hospital if you’re feeling suicidal – you’ll get the holding environment that makes sure you don’t kill yourself for the immediate future, you’ll get connected to a system that can give you useful referrals and medications much faster, and 38% will also end up being directly helped.)

So now what I tell people is the Cliffs’ Notes version of the above – “I’m sorry you have to be here, but we are going to keep you for a few more days to evaluate you, your estimated day of discharge is X but that’s not a promise, if there’s anything specifically making you uncomfortable please let me or the nurses know and we’ll see what we can do.”

I can’t figure out a good way to say the spiel without the last sentence, which is too bad because then they do let me and the nurses know things. Most of them are things that I, as a low-ranking doctor who cannot totally rearrange the unit according to my will, have no ability to change. Some of them are things nobody can change.

Like! It turns out when you lock constitutionally anxious people in a new environment full of psychotic people, they become really really anxious. They tend to request antianxiety drugs. I am happy to give them reasonable doses of the non-addictive anti-anxiety drugs, which then totally fail to do anything, because their idiot outpatient psychiatrist was giving them heroin mixed with horse tranquilizers every day or something. They demand whatever they were getting on the outside, but twice as much, and I can’t give it to them even if I want to because of our safety policies. And now I’m the bad guy.

Or! Some people don’t like noise. I sympathize with this as I am just about the most misophonic person in the world. On the other hand, there’s always one screamer in a psychiatric hospital. Sometimes this screamer chooses to do their thing at four in the morning. The law gives us limited ability to lock them in a soundproof room, and definitely not all the time. So if you are startled by loud noise, you are kind of out of luck. Even if we can put you on the other side of the ward, you’re still going to be bothered by staff coming in your room every fifteen minutes to make sure you haven’t killed yourself, which they are legally required to do. You can complain that the lack of sleep is hurting your recovery, and I believe you, but aside from showing you where we keep the earplugs there’s not much I can do. Once again, now I’m the bad guy.

Add to this people with picky tastes that our kitchen can’t satisfy, people who get bored in the absence of some kind of entertainment we can’t provide, smokers who are unsatisfied by nicotine patches, and the occasional very honest drug addict who just wants some drugs, and I spend about 30% of my day patiently explaining to people why their preferences are totally reasonable and I realize they’re in pain but there’s nothing I can do for them at this moment.

And I know it sounds really selfish of me to say so, but this is really exhausting.

As you may have guessed, I do not very much like inpatient work. You can adjust to having to treat someone having a seizure. You can adjust to somebody banging on the window and screaming. But it’s really hard to adjust to constant moral self-questioning.

IV.

Now I am halfway done with my residency. I will be switching to outpatient work. Everyone who sees me will be there because they want to see me, or at worst because their parents/spouses/children/friends/voices are pressuring them into it. I will be able to continue seeing people for an amount of time long enough that the medications might, in principle, work. It sounds a lot more pleasant.

I have two equal and opposite concerns about outpatient psychiatry. The first is that I might be useless. Like, if someone comes in complaining of depression, then to a first approximation, after a few basic tests and questions to rule out some rarer causes, you give them an SSRI. I have a lot of libertarian friends who think psychiatrists are just a made-up guild who survive because it’s legally impossible for depressed people to give themselves SSRIs without paying them money. There’s some truth to that and I’ve previously joked that some doctors could profitably be replaced by SSRI vending machines.

The second concern is that everybody still screws it up. There’s an old saying: “Doctors bury their mistakes, architects cover theirs with vines, teachers send theirs into politics.” Well, outpatient psychiatrists send their mistakes to inpatient psychiatrists, so as an inpatient psychiatrist I’ve gotten to see a lot of them. Yes, to a first approximation when a person comes in saying they’re depressed you can just do a few basic tests and questions and then give them an SSRI. But the number of cases I’ve seen that end in disaster because their outpatient psychiatrist forgot to do the basic tests and questions, or decided that Adderall was the first-line medication of choice for depression – continues to boggle my mind. So either it’s harder than I think, or I’m surrounded by idiots, or I’m an idiot and don’t know it yet. In which case I’m about to learn.

# The Case Of The Famous Physicist

I.

Old news, but I only just heard about it: Long Island woman says psych ward doctors believed she was delusional for insisting Obama follows her on Twitter.

The story: a woman was brought in for psychiatric evaluation. During the evaluation, she said President Obama followed her on Twitter. The psychiatrists decided she was psychotic and forced medication on her. But in fact, President Obama does follow her on Twitter, just as he does six hundred thousand other people. So they committed a perfectly sane person for telling the truth, leading to what the article calls a “frightening eight-day ‘One Flew Over the Cuckoo’s Nest’ ordeal”.

I don’t know anything about this case or this person, and I definitely don’t want this to sound like I have anything to say one way or the other about this person I have never met. But I’ve been involved in enough similar cases to have a different perspective, and wonder whether it was quite as much of an outrage as the article makes it out to be.

Consider: the comment occurred when she was in a hospital for psychiatric evaluation; that is, she was brought in before the Obama comment. According to the article:

The bizarre experience began Sept. 12, when the NYPD seized her prized 2003 BMW 325Ci in Harlem because they suspected she was high on weed, her attorney, Michael Lamonsoff, said. Cops found no marijuana but confiscated her ride anyway, he said. The NYPD declined to comment.

The following day, Brock walked into the NYPD’s Public Service Area 6 stationhouse in Harlem to retrieve her car, her suit charges. Brock — an eccentric 32-year-old born in Jamaica with dreams of making it big in the entertainment business — admitted in an interview she was “emotional,” but insisted she in no way is an “emotionally disturbed person.” Nevertheless, cops cuffed her and put her in an ambulance bound for the hospital, her suit charges.

This sort of elides over everything in between “went into police station to ask for car back” and “cops put her in an ambulance bound for the hospital.”

I’ve had patients sent by ambulance from the police station. It’s almost always because they started screaming and yelling threats at the police. Now, screaming and yelling threats at the police, although not a very good idea, is not always evidence of psychiatric disorder. But it often is. If you’re manic or on drugs, you’re a lot more likely to have the particular type of bad judgment that makes screaming threats at the police a seem like a good idea.

I don’t know what happened with Ms. Brock, and the article doesn’t say. I did take a look at her Twitter account, which is mostly angry tweets about Trayvon Martin, Mike Brown, and police brutality. A typical example is “Police should NOT be allowed to Murder the Citizens of this Country!!! We are not Animals!!! @BarackObama fix it NOW!!! #BanGuns”. Other Tweets seem maybe a little threatening, like the ominous “@NRA you people deserve the wrath that will come on you in the very near future” and “#KillRacists”.

So – cops have just done something very unfair to a person who likes making threats and doesn’t like cops, and who herself admits to being “eccentric” and “emotional”. Again without knowing what went on, my guess was that her “discussion” with the police was sufficiently exciting that they thought calling an ambulance for a psych evaluation was in order.

So imagine you’re the psychiatrist, you’re receiving a patient from the police for evaluation after she got in some kind of screaming match with them. And now she tells you Barack Obama follows her on Twitter. The article says:

“I told (the doctor) Obama follows me on Twitter to show her the type of person I am. I’m a good person, a positive person. Obama follows positive people!”

This is weird. At best, it displays a surprising ignorance of how Twitter and/or the world works. Yes, Obama follows 600,000 people on Twitter, but this does not prove that each of them is “a positive person”. I would assume he has some program that auto-follows anyone who mentions him. When you think this is a good thing to bring up during, of all times, a psychiatric evaluation, then I start to wonder.

(I should also add here that, in my limited experience, social media is God’s gift to grandiose psychiatric patients. None of them are “a guy with a Facebook page”. They’re all “social media celebrities with hundreds of followers”. It’s always “YOU CAN’T DO THIS TO ME! DO YOU KNOW WHO I AM? I HAVE HUNDREDS OF FOLLOWERS ON TWITTER! EVEN [NAME OF TWITTER PERSON I HAVE NEVER HEARD OF] FOLLOWS ME! THIS IS GOING TO GO VIRAL!” One patient even told me, in a threatening manner, that his blog had over a thousand hits. “You mean a day?” I asked. “No, total,” he answered. Then he wondered why I was so utterly failing to look impressed.)

So I’ll be honest – if someone had just been brought in for a shouting match with police, and the first thing they told me was that President Obama follows them on Twitter, well – among the most common symptoms of mania are irritability, grandiosity, poor decision-making, and flamboyant dress/behavior. Getting in a fight with the police sounds like irritability and poor decision making, thinking the Obama follow is relevant sounds like grandiosity, and this woman’s Facebook photo certainly suggests a flair for the dramatic. I’m not saying I would commit her on these factors alone, and a lot would depend on the rest of the history. But she already would have started digging quite a hole for herself.

Remember, psychiatrists have to err on the side of admitting people. Even if this lady didn’t have bipolar disorder, admitting her for further evaluation seems like the sort of thing that could be an honest mistake.

(and remember, all we have here is her side of the story. Goodness only knows what other things she might have said or done that she didn’t leak to the news for their sympathetic article based entirely on her testimony.)

But then there’s the hospital’s subsequent conduct. They said that she was unemployed. She said she was employed and could prove it through her Internet presence. The hospital apparently didn’t check and continued to say she was psychotic for thinking she was employed. How could the hospital possibly be so incompetent as to not check the link to her employment documents that she was personally giving them?

II.

Let me tell you about Professor T.

(this would be a good time to reiterate that every patient story I tell here is a composite of multiple different people with all of the details changed around to protect anonymity. The gist of the story points out a true thing, but the specifics are all twisted around so thoroughly that even the people involved couldn’t recognize themselves.)

Professor T came to me in handcuffs. The police had picked him up in response to a call at the local university, where he was trying to give a lecture to a class that wasn’t expecting him and didn’t want him. The class’s actual professor had asked what the heck he was doing, and he had explained that he was Professor T, world-famous physicist, and that the head of the college had invited him to give a lecture on his theories that day, and it was too bad she hadn’t communicated this with the rank-and-file teachers, but he was a very busy man and they should all be honored by his presence and stop what they were doing and listen to him. A quick call to the administration confirmed none of them had any idea who he was either, and when he refused to go away or stop trying to lecture, the police were called to remove him. He started yelling and screaming at the police and telling them they were fools who were too small-minded to recognize a great scientist when they saw one, and they’d get what was coming to them. The cops decided this was a job for a psychiatrist and brought him to me.

Professor T said he’d just been really angry that the bureaucracy had screwed up badly enough to make him miss his lecture and that no one was willing to accommodate him. He was, after all, a very important leading scientist with a busy schedule. He moved in elite circles! Famous people like Edward Witten knew him well, and I was welcome to call to confirm that! He was used to being shown more respect!

I got a weird vibe from Professor T during our discussion. I know that distinguishing between professors and hobos is a famously difficult problem, but he just struck me as a little too much towards the hobo end for comfort. So I asked him if I could see any proof that he was who he said that he was.

He was happy to comply, and once we got to a computer he showed me some scientific papers with his name on top, suitably peppered with complicated words like “tachyon” and “chromodynamics”. He showed me a picture of him winning some prestigious physics prize, dressed in a lab coat with a medal around his neck. He even showed me what looked like a press release: “Professor T comes up with new Theory Of Everything that may explain Higgs Boson”. It was pretty convincing.

But.

A bunch of patients had cancelled on me that day, so I had way too much free time. I started looking into Professor T’s credentials a little more. His papers were hosted on a private site and didn’t show up on Google Scholar and didn’t seem to be affiliated with any journal. The press release was on the same website, and seemed suspiciously badly written. There was nobody else in his photograph, and it was impossible to see what was written on the medal. Was it possible that a crazy hobo had just written some things that looked like papers, written something that looked like a press release, and then bought a lab coat and medal and taken a picture with them? Do real theoretical physicists even wear lab coats?

So I asked him if I could talk to Dr. Witten, whom he said he knew well. Professor T agreed. I Googled his phone number and called him up.

(It wasn’t actually Dr. Witten whom I called, but the case I’m adapting this story from involved someone else about as famous)

His secretary answered the phone, and I said I was a psychiatrist, and I asked if I could speak to Dr. Witten. The secretary was reluctant, but when I said it was about Professor T, she immediately asked me to hold, and I got Dr. Witten himself. I asked him if he knew Professor T.

“Absolutely,” said Dr. Witten. “He’s a crazy guy who keeps calling me up and telling me he’s solved physics. I don’t think he’s actually a professor of anything. I read one of his papers once, just for kicks, and it’s just a bunch of science terms like ‘tachyon’ and ‘chromodynamics’ strung together without rhyme or reason. It might fool a layman, but trust me, it makes no sense. I told him to stop calling me, and he wouldn’t, and finally I had to block him on my phone, and now he’s sending me letters in the mail, and it’s always same ranting about tachyon chromodynamics, which isn’t even a real thing. Did you say you’re a psychiatrist? Perfect, I’m so glad he’s finally getting treatment.”

I told Professor T about this, and he nodded his head. “Yes,” he said “I told you that Dr. Witten knows me well. I didn’t say he liked me. He still doesn’t fully understand my theories. But I am sure he’ll come around.”

I kept Professor T in hospital for about a week, and I can’t count how many times he yelled at me and complained that I was being unfair to him by not doing whatever the heck he wanted me to do that day. Read another one of his papers that would convince me his theories were sound. Call up yet another famous physicist he “knew”. Look at yet another of his fake websites devoted to himself. Every day, he threatened to sue me and my boss and the entire hospital for keeping him there even though he’d “proven” to us he was who he said.

Remember, delusions are fixed false beliefs. People are quite sure they’re true, quite sure they have evidence for them, and nothing (except occasionally really good psychiatric treatment) will convince them otherwise. They’ll keep demanding you take time to investigate more and more bizarre “arguments” and “evidence”, and if you ever stop, even after days and days of everything they say being one hundred percent refuted, they’ll accuse you of acting in bad faith.

(it’s like Internet arguments, only more so)

In everyday life, we get by on an assumption of trust. If I tell my boss I’m sick, he probably believes me. If he doesn’t believe me, and I send him a doctor’s note, he probably believes that the doctor’s note isn’t forged. If he doesn’t believe that, and he asks me for a number to call the doctor at, he probably believes it’s a real doctor and not my brother pretending to be a doctor to help me out. Yes, there are a couple of people who abuse that trust, but few enough that the rest of us are usually able to get by.

In psychiatry, there are a bunch of delusional people, paranoid people, narcissists, compulsive liars, and others who deliberately or unknowingly stretch the truth past the breaking point. Worse, a lot of the cues we use to detect liars, like “Are they shifty-looking?” don’t work, either because the person involved really believes what they’re saying or because they’re too far from the neurotypical norm for our usual intuitions to apply. A lot of the assumptions of trust we usually use crash and burn. If the person sitting next to me on the train says he’s a physics professor, I believe him. If the person brought in by police for a psychiatric evaluation says he’s a physics professor, maybe I don’t, and “how much time do I spent assessing the evidence and how much do I believe?” is a really tough question.

I am not trained as a police officer, detective, or judge. I’m also not paid to do their jobs. I’m also stuck in a system where the primary incentive is that if I ever fail to commit someone, then if they do anything bad after that I can be sued for everything I own. So I am stuck drawing partial conclusions, from incomplete evidence, in time I don’t have, from people I can’t necessarily trust, without even the ability to err on the side of caution.

I don’t think the hospital in the article followed great practices – in particular I’m unclear on how they came to believe the person was unemployed. And sending her the bill for her own involuntary commitment is an obvious injustice (albeit a universally practiced one). But the mistakes in the admission process are all ones I can imagine any psychiatrist making. Including me.

And that means something. You can trust me. After all, four different Dalai Lama accounts follow me on Twitter.

# The Argument From Cultural Evolution

[Content note: Discussion of debate over gay marriage]  
[Epistemic status: Very preliminary. Probably missing something. Looking forward to hearing what.]

The term “cultural evolution” is getting used a lot nowadays. In its simplest form, it just means cultures preserve useful ideas and tips. For example, as per Carcinization:

[The Lost European Explorer] experiment has been repeated many times when European explorers were stranded in an unfamiliar habitat. Despite desperate efforts and ample learning time, these hardy men and women suffered or died because they lacked crucial information about how to adapt to the habitat. The Franklin Expedition of 1846 illustrates this point. Sir John Franklin, a Fellow of the Royal Society and an experienced Arctic traveler, set out to find the Northwest Passage, and spent two ice-bound winters in the Arctic, the second on King William Island. Everyone eventually perished from starvation and scurvy. The Central Inuit have lived around King William Island for at least 700 years. This area is rich in animal resources. Nonetheless, the British explorers starved because they did not have the necessary local knowledge, and despite being endowed with the same cognitive abilities as the Inuit, and having two years to use these abilities, failed to learn the skills necessary to subsist in this habitat.

Presumably, the Inuit neither conducted deliberate centralized experiments to determine what food in their area was edible, nor derived the information from explicit understanding of the principles of nutrition. Rather, over thousands of years, various proposals like “eat those yummy-looking red berries that grow on the small bushes” and “always hunt seals in large groups” were accidentally tested, with the successful ones spreading until they became universal tradition and the disastrous ones being warned against as taboo. Without any deliberate effort, the Inuit ended out with a remarkably effective set of survival techniques.

Something like this seems so obviously true as to not require further discussion.

However, recently “cultural evolution” has slipped, without much consideration, into a much stronger meaning. For example, in his commentary on Ross Douthat’s article on gay marriage, Tumblr user severnayazemlya writes:

What Douthat is saying is that there was some system that existed sometime in the past that was more human-shaped than Marcotte’s vision for the future. Gavin McInnes has said the same.

The conservative argument is that the cultural inheritance that the past hands down to the present is more human-shaped than most reforms proposed in the present – because there were reformers in the past, and, absent major breaks in the continuity, past reforms have had time to be tested for their fit: those that worked were kept, and those that didn’t were discarded.

Likewise, on one of my recent blog posts, commenter Steve Johnson writes:

Every surviving cultural tradition on Earth is hostile to homosexuality – that’s no accident. That’s cultural evolution in action.

This form of cultural evolution seems to work something like so: our culture, and indeed most cultures, used to have a certain conception of marriage. That conception of marriage outcompeted other conceptions of marriage from the distant past all the way to the present. Societies with alternative conceptions of marriage seem to have died out. That suggests that this conception contains something useful; even if we can’t see it we should be wary of interfering with it, in the same way we are wary to disrupt our body’s metabolic balance or alter genes willy-nilly.

The difference between the obvious Inuit form of cultural evolution and the non-obvious marriage form is that of within-culture versus between-cultures evolution.

Consider: one Inuit tries the red berries and discovers they make her sick. Out of pure self-interest, she decides not to eat them again, and tells her friends the same. Also out of self-interest, they decide not to eat them; those who think they can get away with eating them anyway are quickly disabused of the notion. The taboo against eating red berries quickly spreads throughout the culture.

Marriage doesn’t seem to work that way. If one person decides not to marry in the usual way, it doesn’t necessarily hurt that person. They might have lots of affairs, and enjoy them. Or they might get gay married, and enjoy that. Any claim that cultural evolution argues against gay marriage because it’s bad for the actual gay-married person must face the fact that actually gay-married people seem totally okay with it, and in fact are urging their friends to do it, the exact opposite of the red berry situation.

So I interpret it as a different claim: a culture that allows gay marriage will, for various reasons, become weak and unsuccessful. Then it will be crushed by other cultures, either militarily, economically, or in a sort of marketplace of ideas where people convert to or assimilate into the other culture because it’s more attractive and successful.

Note that THIS IS REALLY DIFFERENT FROM THE FIRST TYPE OF CULTURAL EVOLUTION. In fact, it might be diametrically opposite. For example, gay sex may be lots of fun – and as people figure this out and tell their friends, it will be positively selected through the first type of cultural evolution. But it might weaken a culture’s Moral Fabric – in which case it will be negatively selected through the second type of cultural evolution.

This is sort of group selectionism, but in this case I’m okay with it. Consider the analogy of a cancer cell. Becoming cancerous makes a cell much more likely to spread within its organism – the equivalent of positive intracultural selection – but also makes its organism at a severe disadvantage compared to other organisms – the equivalent of negative intercultural selection. As a result, we expect organisms to evolve strong internal defenses against cancer – which in fact they have. In the same way, it’s plausible that cultures might evolve strong internal defenses against actions that are fun but Weaken Moral Fabric, and sure enough we find that everything halfway enjoyable comes with a lecture from our elders about Why We Shouldn’t Do It. Presumably if these things really Weaken Moral Fabric in an important way, then those cultures that develop strong internal defenses against them – for example, a strong and well-enforced religious taboo against gay marriage – will be more likely succeed while other cultures die out.

So in principle this kind of intercultural selection could happen. In practice I think the effect is negligible.

Evolutionary biology has a lot of equations to calculate how long it will take a positively-selected trait to spread. It depends on a lot of different things, but the most salient here are the length of a generation for the affected organism, and the extent of the selective advantage conferred by the trait.

How long is a “generation” in cultural evolution? Rome lasted a thousand years, Byzantium another thousand. It took about three hundred years for Christianity to replace paganism in Rome; Enlightenment values have been replacing Christianity for three hundred years already and aren’t nearly done. Any sort of evolutionary process that involves waiting for Rome to fall is a process that will take way longer than human history to come to any sort of conclusion.

How much advantage can an individual cultural trait confer? Probably very small in the grand scheme of things. Compare Judeo-Christian attitudes about sex to Greco-Roman attitudes about sex. One might argue that the Judeo-Christian attitudes are superior, since Christianity did eventually take over Rome. On the other hand, both Greece and Rome took over Israel at various points; various Jewish texts record that during that time a lot of Jews were defecting to Greco-Roman culture and there were precious few defections the other way. It would seem that all of the other differences between Greco-Roman and Judeo-Christian culture – theology, non-sexual mores, geography, technology, philosophy – had a lot more effect than the sexual mores.

This is also unsurprising from a population biology perspective. Suppose that one of my children gets a mutation causing 1% less risk of infectious disease. This is fine, but they might get killed in a car accident before their tenth birthday, or be too ugly to find a partner, or get an infectious disease anyway because 1% less risk isn’t really much less risk. If my child survives, and passes her mutation on to millions of other people all with their randomly distributed level of other good and bad genes and good and bad luck, then maybe eventually over thousands of generations, people with the new beneficial mutation will take over from people without it.

But, as mentioned above, we don’t have thousands of generations for cultural evolution to do anything. Communism, which basically took all of the worst ideas in history, combined them together into a package deal, and said “Let’s do all of these at once”, took almost a century to collapse, and still hasn’t collapsed in a couple of places. Imagine if instead of Communism happening, twenty different countries had adopted one Communist ideal each and we’d waited to see which ones grew and which ones declined. We’d still be waiting, and probably instead of getting any useful information we’d just end up seeing the Rise of China and not being sure whether it was because of their Communist ideal or something else.

The closest thing I can think of to anyone actually gaining useful information out of cultural evolution is the failure of various small communes and social experiments. But once again, these only failed because they tried all the bad ideas at once, and a big part of their failure was intracultural evolution – the people involved noticed they, personally, were poor and unhappy, said “screw this”, and went back to non-communal living.

So overall, I think any appeal to intercultural evolution as having proven anything is on very shaky ground. Appeal to intracultural evolution is much more reasonable, but crucially, can’t be used to override people’s own decisions about whether they’re happy doing something or not. If someone says “I enjoy this, and I’ve been doing it a few years and not noticed any bad consequences, and I suggest you do it as well”, then you’re going to have a hard time arguing against the practice on grounds of cultural evolution.

[EDIT: Actually, this leaves out a possible third kind of cultural evolution, where cultures try good ideas, learn to like them, and stick with them; or try bad ideas, learn to hate them, and stop. For example, China experimented with Maoism, that didn’t work, then experimented with a variant of capitalism and liked it enough to stick with it thus far. Given that something similar is happening on smaller scales (eg experimentation in policing, education, budgeting, etc) all the time, you could eventually end up with a pretty finely-tuned culture. This obviously happens, but it seems loaded to think of this as ‘cultural evolution’ instead of just ‘guess and check’ or ‘learn from history’. The former formulation suggests something illegible to understanding; the latter formulation suggests that it happens by deliberate human responses to bad consequences, and makes it less of an excuse for general conservativism.

If you say “Let’s be Maoist!” I can say “No, the Chinese tried that, it led to X, Y, and Z consequences, and then they switched to something else and things got better.” The cultural evolution argument for traditional sexuality seems to be trying to argue in the absence of, or in parallel to, such observable historical lessons.

Likewise, there’s a cultural evolution argument that we tried traditional sexuality, that made a lot of people unhappy, and now we’re trying something else. It’s unclear how this is different from the Maoism example in a way that makes jettisoning Maoism good, but jettisoning traditional sexuality bad.]

# Cultural Evolution 2: Thanks For The Meme-Rise

Some points I gleaned from the comments of yesterday’s post:

1. Cultural evolution can happen in cases where a super-innovation allows one culture to conquer or overwhelm all others. For example, agricultural groups were (after a long transition period) eventually able to overwhelm hunter-gatherer groups, even thought for an individual hunting-gathering was probably more enjoyable than agriculture. Likewise, industrialized societies were pretty quickly able to outcompete nonindustrialized societies, and either colonized them or forced them to industrialize in turn to keep up. Both of these seem like clear-cut examples of cultural evolution. But they only work because of a really big fitness advantage; industrial societies are on a whole other level from preindustrialized ones. It doesn’t necessarily generalize to saying that small, moderately beneficial ideas will catch on, or slightly detrimental ones be selected against.

2. Cultural evolution can happen when one group in a society outbreeds another. The Amish population has increased twenty times faster than the non-Amish American population in the past century. At a constant growth rate, it’ll be only another four hundred years or so before America is an Amish-majority nation. More seriously, some people expect something like this to happen with high-fertility-rate immigrant populations, like Latinos and Muslims. In cases of strong differential fertility rates, cultural evolution becomes a race to see if the faster-growing minority can reproduce faster than the majority can assimilate them. However, despite dire predictions of all of us being crushed under the Amish’s quaint hand-made boots, people had trouble thinking of historical examples of something like this happening. Sure, populations have replaced other populations – like the Anglo-Saxons replacing the Celts in England – but it’s tended to occur alongside military invasions.

3. Cultural evolution can happen with units smaller than Rome-sized grand civilizations. Several people brought up subcultures, like hipsters and Goths, and noted that these have “generations” on the order of a few decades, and thus could potentially undergo evolution conforming to population genetic equations in a reasonable amount of time. Because they’re larger units than just a single person, their “evolution” could select for things that bind people together, like rituals and cohesion-building symbology and so on, and be more interesting than just individual memetics. They could also spread very quickly as people rush to join the attractive ones. Okay. But subcultures like Goths seem like a very modern phenomenon, and I can’t think of ancient examples of, for example, a subculture that became popular and spread and became dominant/universal. Religions are the closest thing here, but they have lifespans measured in centuries and don’t seem to be a big improvement over waiting for the Fall of Rome.

4. Cultural evolution can occur by an accretion of things that work. For example, the first rituals might have been impromptu celebrations of specific events, but because they helped people bond, people kept doing them. But this seems to require some human intelligence to notice “Hey, we seem to be bonding better ever since we implemented that ritual, let’s keep doing it”. Without that, it collapses back to the sort of intercultural evolution where the culture is 1% better and after thousands of cultural generations lasting millennia each it outcompetes others. That makes it unsatisfying for people who want to use cultural evolution as a grounding for Chesterton’s Fence, ie “we don’t know why we do this, but we ought to keep on doing it.”

5. Cultural evolution could have occurred way way back in prehistory. There seem to be about 50,000 years of prehistory, there were many more cultures back then, and maybe cultural generations were shorter – for all anybody knows, clans could have disintegrated and reformed over the space of decades. That provides enough generation time for cultural evolution to work. Question is, can we trust anything that evolved in pre-history – when the pressing social issues of the day were things like “How do we not get eaten by bears?” – to still be relevant?

There does seem to be the potential for cultural evolution to be interesting, but I’m still not seeing it as a strong argument for preserving particular features of inherited culture absent other arguments suggesting we know why we want those things to be preserved.

[Edit: An alternative ontology]

# Things That Sometimes Work If You Have Anxiety

Anxiety disorders are the most common class of psychiatric disorders. Their US prevalence is about 20%. They’re also among the least recognized and least treated. We have sort of finally beaten into people’s thick skulls that depression isn’t just being sad, and you can’t just turn your frown upside down or something – but the most common response to anxiety disorders is still “Anxiety? So what, everyone gets that sometimes.”

But it’s hard to describe how disabling anxiety can be. A lot of people with nominally much worse conditions – depression, bipolar, even psychosis – will insist that they want their anxiety treated before anything else, because they can live with the rest. On the other hand, while a lot of people with psychosis have enough other problems that treating the psychosis barely puts a dent in their issues, a lot of people with anxiety would be happy and productive if they could just do something about it.

Since I’ve gotten some positive comments on my discussion of depression treatments I thought I’d go through some of the things I’ve seen used to treat anxiety. I’ll include the same disclaimer:

This will be inferior to reading official suggestions, but you will probably not read official suggestions, and you may read this. All opinions here are my own, they are not endorsed by the hospital I work at, they do not constitute medical advice, I have a known habit of being too intrigued by extremely weird experimental ideas for my own good, and you read this at your own risk. I am still a resident (new doctor) and my knowledge is still very slim compared to more experienced professionals. Overall this is more of a starting point for your own research rather than something I would expect people to have good results following exactly as written.

I’ll mostly be talking about what’s called generalized anxiety disorder, with some applicability to panic disorder. Social anxiety, specific phobias, et cetera are their own thing, as is anxiety secondary to other illnesses – but some of the advice may cross over. I’m not going to get too into diagnosis, because generalized anxiety disorder is pretty much exactly what you think it is and a lot (though not all) of this will be applicable for subclinical anxiety as well.

I. Diet And Lifestyle

You didn’t think you were going to get out of this part, did you?

Pretty much every study – epidemiological or experimental, short-term or long-term, has shown that exercise decreases anxiety. The effect seems limited to aerobic exercise like walking, running or swimming, preferably for longer than twenty minutes. Various mechanisms have been postulated including norepinephrine, endogenous opioids, and decreased inflammation.

There’s less agreement on diet. The people who hate fat says high-fat diets cause anxiety. The people who hate carbs say high-carb diets cause anxiety. The people who hate processed food say processed foods cause anxiety. The people who recommend fish oil for everything say insufficient fish oil causes anxiety. None of it seems super credible, but Mayo Clinic has some suitably bland advice.

The one very important connection – if you drink too much coffee, or any other source of caffeine, that will make you anxious. I once had a patient come to me with severe recurrent anxiety. I asked her how much coffee she drank, and she said about twenty cups per day. Suffice it to say this was not a Dr. House-caliber medical mystery.

Also needless to say: get enough sleep. Seriously. Get enough sleep.

Many people find that various breathing exercises or other sorts of mindfulness activities can be helpful in the short term and sometimes build skills useful for the long term. My hospital gives people these handouts on breathing techniques and progressive muscle relaxation. I’ve made fun of HeartMath in the past, but I only learned about them because many people find some success, probably placebo-ish, with their quick coherence technique. If you’re an overachiever and want to get really into this sort of stuff, people always say good things about yoga and especially pranayama breathing. Studies seem to back this up (1, 2, 3) though you’ve got to be careful to weed out the studies by very religious Hindus trying to prove they’ve been right all along.

Meditation has similarly positive results. Here’s a study showing that an intervention to teach patients meditation resulted in decreased anxiety with p < 0.001 even three years later. Here's a meta-analysis of 39 studies finding an effect size of about 0.6 (medium) in the general population, and an effect size of about 1.0 (large) in people with anxiety disorders. But here’s an equal and opposite review that found only “equivocal” results. As far as I can tell, most people investigating meditation think it works pretty well. The meditation techniques that seem to work best are mindfulness meditation and transcendental meditation. You can learn a little about mindfulness meditation here. In order to learn about Transcendental Meditation, send a check made out for $5000 to Maharishi Mahesh Yogi, PO Box….

II. Therapy

Cognitive-behavioral therapy works okay for anxiety just like it works okay for everything else. The Big Graph O’ Effect Sizes says that psychotherapy on average has an effect size of 0.51 in generalized anxiety, compared to medication’s 0.31. This shouldn’t be taken too seriously – the confidence intervals overlap and there’s a wide range of efficacy for different medications – but you won’t be doing any worse by going for the therapy first. Even the Cochrane Review, famous for never drawing any conclusion other than “more research is needed”, is tentatively willing to say that psychotherapy works for anxiety disorders. Their study trends towards finding that cognitive behavioral therapy works better than supportive therapy, but is unable to prove significance – apparently more research is needed.

Exposure therapy can also be useful for panic attacks or specific phobias. This is where they expose you to the thing you’re scared of (or deliberately initiate a panic attack) and keep doing it until you stop being scared and start being bored. According to a bunch of studies it works neither better nor worse than cognitive-behvioral therapy for most things, but my unsupported impression has always been that it’s better at least for panic disorder. Cognitive-behavioral therapy seems clearly superior for social phobia.

You can get psychotherapy from any qualified psychotherapist, a category including counselors, social workers, psychologists, and sometimes psychiatrists. Ones who use “a school” (for example, describe themselves as practicing cognitive behavioral therapy) are usually considered better than those who don’t (“Oh, I do a little of everything with every patient”). If you can’t find (or don’t want to find) a good therapist, there is preliminary evidence that a good self-help therapy workbook (“bibliotherapy”) is about as good as real therapy – including for anxiety (study, other study, yet another study).

I have no special insight into which self-help workbooks are any good, but The Cognitive Behavioral Workbook for Anxiety: A Step-By-Step Program seems to get pretty good ratings.

III. Medications

To be tried after diet and lifestyle interventions when possible.

Medication can work either instead of or in addition to therapy. There are at least seven categories of commonly used conventional anxiety medications: SSRIs, SNRIs, antihistamines, antipsychotics, anticonvulsants, benzodiazepines, and azapirones. These can be divided into mostly-acute (antihistamines and benzos) and mostly-long-term (SSRIs, SNRIs, anticonvulsants, azathioprines), with antipsychotics kind of being a tossup. Depending on whether you just need to get through the occasional panic attack or whether you’re in a chronic unremitting anxiety state, you might want one, the other, or both.

You probably know antihistamines (example: Benadryl) from the many common over-the-counter members of this class. They have some mild short-term anti-anxiety effects. Benadryl will work in a pinch if you need something without a prescription, but the most commonly used anxiolytic antihistamine is hydroxyzine (“Vistaril”, “Atarax”), which is a bit more powerful and less likely to make you fall asleep. As far as anxiolytics go it’s pretty safe as long as it doesn’t make you too sleepy. If you just need something to take the edge off the occasional anxiety attack, this works fine.

Benzodiazepines (examples: Xanax, Ativan, Valium, Klonopin) are very effective in the short-term but also very controversial. In some people they are very habit-forming and can produce a picture very similar to addiction to alcohol (which they chemically resemble). Keep in mind how bad an idea it might be to become extremely addicted to prescription pills that you may suddenly lose access to depending on how your doctor is feeling (you might expect doctors would take the difficulty of coming off these drugs into account, but you might expect a lot of things from doctors that don’t always happen). Studies suggest benzodiazepines can sometimes build tolerance, and that after a month or two of frequent use, they lose their positive effect and you need them just to feel normal. That having been said, a subset of patients – and I can’t tell at this point if it’s a majority or a minority – go on benzodiazepines, do very well, stay on them for long periods without getting dependent, and never have anxiety again. It’s kind of a crapshoot. The most generally recognized “safe” use of benzos is the occasional Xanax to deal with rare but very stressful situations (for example, flying on an airplane if you’re scared of heights). Other people say Klonopin is safer than some of the others and that it’s worth a shot as long as you realize that “Klonopin dose gradually creeping upwards” is a sign that you’re getting into a bad place and need to react immediately. Most people recommend trying other things first before you come here, but once you’ve exhausted other options these can be a powerful last resort.

SSRIs (examples: Prozac, Celexa, Lexapro, Zoloft) are the mainstay of chronic anxiety treatment just like they’re the mainstay of chronic everything-else treatment. As usual, they have real but modest effects after about a month or so, more in some people and less in others. As usual, if one SSRI doesn’t work for you, you might want to try another. These are pretty safe aside from the sexual side effects. Some people get mild withdrawals if they go off these too quickly, so don’t do that. A lot of people use both an SSRI for chronic treatment, plus either an antihistamine or benzo for “break-through” anxiety.

SNRIs (examples: Effexor, Cymbalta) are like SSRIs, but for two neurotransmitters instead of one. This is supposed to make them a little bit more effective. Maybe they are, maybe they aren’t. Fewer sexual problems than SSRIs, but worse discontinuation syndrome. They’re a good second-line chronic medication if SSRIs don’t work. Effexor is probably the best.

Azapirones (example: BuSpar) is, unusually, a rare drug which is specifically targeted at anxiety, rather than a being a repurposed antidepressant or something. BuSpar is very safe, not at all addictive, and rarely works. Every so often somebody comes out with a very cheerful study saying something like “Buspar just as effective as benzodiazepines if given correctly!” and everybody laughs hysterically and goes back to never thinking about it.

Anticonvulsants (examples: Depakote, Neurontin, Tegretol, Lyrica) are seizure medications that sometimes sort of work for anxiety. Most of them have strong side effects and limited utility. The exception is Lyrica (pregabalin), which is pretty new but has shown excellent safety and efficacy in studies. It doesn’t have an FDA indication for anxiety and it’s pretty expensive, so you might have a hard time getting it, but it is at least a well-kept secret.

Atypical antipsychotics (examples: Seroquel, Zyprexa, Abilify, Geodon) are, as always, overused. Most of them either make you gain lots of weight, put you at increased risk for heart rhythm problems, make you feel terrible, put you at risk of permanent movement disorders, or all of the above. They do often treat anxiety, sometimes very well, and psychiatrists like them because they’re good all-purpose no-nonsense drugs with big advertising budgets, but unless you’re also psychotic consider trying some other things first before you try these.

An article in Journal of Psychopharmacology tries to compare the efficacy of all of these classes of drugs and gets the following effect sizes (bigger number = bigger effect):

Pregabalin: 0.5  
Antihistamine: 0.45  
SNRI: 0.42  
Benzo: 0.38  
SSRI: 0.36  
Azapirone: 0.17  
Alternative medicine: -0.31

(remember, other studies suggest psychotherapy is around 0.5)

I heavily challenge the claim that antihistamines are more effect than (or anywhere near as effective as) benzos. I don’t know the confidence intervals on these numbers, so I would suggest reading it as “Everything is about equally effective, except azapirones which aren’t as good”. Their “alternative medicine” category was mostly kava and homeopathy, and I have no idea why it came out negative (kava’s pretty good, and homeopathy shouldn’t separate from 0).

There are also some less commonly used drugs that might help people who don’t respond to any of these.

As usual, MAOIs are very effective, moderately dangerous, and super hard to get. They seem to work especially well for panic disorder and social anxiety.

Clonidine is a medication usually used to control blood pressure. It’s somewhat effective against anxiety and some people think it should be used more. But it can cause you to become too sedated (abnormally low heart rate) and in some people it makes anxiety worse for some reason.

Beta-blockers (example: propranalol) are another blood pressure medication. It is especially effective against somatic symptoms of anxiety – racing heartbeat, shaking, et cetera – and sometimes getting rid of those can make the anxiety go away entirely. It’s most famous for its use against performance anxiety: about a third of musicians use them in concerts, and I’ve heard similar rumors about public speakers, actors, et cetera. I used to think this was a little-known piece of trivia, but whenever I bring it up to doctors (“Hey, did you know some people use beta-blockers for performance anxiety”) the usual response is “Oh, yeah, I prescribe myself some of that when I have to give a presentation at grand rounds.” They don’t seem quite as good for longer-term anxiety disorders, though some people have had good results with them.

I once saw an excellent psychiatrist whom I deeply respect try everything on a patient with severe treatment-resistant anxiety with no results whatsoever until finally he came to Thorazine. This treated the patient’s anxiety pretty well, at the cost of provoking quite a bit of anxiety in the doctor.

Without meaning to give medical advice, and with the caveat that you should ask your doctor for their opinion – one good pharmacological treatment algorithm for anxiety disorders is:

If you just have occasional outbursts that bother you, take occasional doses of hydroxyzine.

If you have a longer-term problem, start with an SSRI. If that doesn’t work, either try more SSRIs and SNRIs, or go to Lyrica. You might as well be on BuSpar somewhere in the process too. If none of that works, choose your poison (or have it chosen for you) among MAOIs, benzos, clonidine, or antipsychotics.

IV. Alternative Treatments

To be used out of curiosity or desperation only – you have other options and these are not guaranteed safe or effective.

Massage therapy, acupuncture, aromatherapy, and everything else in the category of “unnecessarily medicalized relaxing thing” all perform very well as long as you don’t look too hard for a suitable control group. Yes, these are probably placebo, but they’re very effective placebos and if they both work I would rather take a placebo than an antipsychotic.

Inositol and l-theanine are both found in small quantities in the diet (inositol in some vegetables, theanine in tea) and supplementing them has been inconsistently found to help with anxiety. Inositol had some preliminary evidence for effectiveness in panic disorder, but a more recent meta-analysis was unimpressive. I can only say that I have some anecdotal evidence of extremely positive reactions to inositol, but we all know what they say about anecdotal evidence. Keep in mind that the dose used in studies is way larger than the dose anyone will give you – usually corresponding to about 20 of those 500 mg inositol pills a day. This makes it expensive and inconvenient, and most people just compromise by taking so little inositol it shouldn’t possibly be able to have any effect. L-theanine also has a lot of small studies in support, although there’s some question on whether it works on its own or whether it just has useful synergistic effects with caffeine. Sun-theanine is generally considered the most effective form, and recommended dose is about 100 – 400 mg. Both these supplements are afaik very safe and a good option for people who want to test things that might or might not work but have minimal risk. Magnesium should also be in here somewhere.

GABA is the main inhibitory neurotransmitter in the nervous system, and a lot of these other interventions are attempts to convince the brain to release more GABA or potentiate the GABA that’s already released. Can we just cut out the middleman and ingest GABA pills directly? The supplement industry would like you to think so, and you can certainly buy them anywhere supplements are sold, but it’s generally believed that orally ingested GABA can’t cross the blood-brain barrier. The Russians have developed a modified version of GABA that doesn’t have this problem; called picamilon, it seems to be a pretty popular anxiety treatment on the other side of the Pharmacological Iron Curtain. It’s pretty easy to get as a non-prescription supplement here in the West. There are very few studies on it, the ones that exist are in Russian, and I have nothing to go on but a couple of anecdotal reports, most of which are positive (though I personally noticed no effects). But the mechanism of action is plausible, and the long history of successful Russian use at least suggests it probably won’t kill you immediately. Most common dosage seems to be about 100 – 300 mg.

The nootropics/supplement/nutraceutical community also suggest ashwagandha and bacopa for anxiety; various low-quality studies support the use of both (ashwagandha meta-analysis, bacopa study 1, bacopa study 2, bacopa study 3). Bacopa may take several months of frequent use before it starts working; I tried it briefly and had to stop because of gastrointestinal side effects, which are pretty common. There’s also some worry around heavy metal contamination. Swanson’s and Nootropic Depot’s are two that have third-party testing showing they’re uncontaminated.

Kava is a traditional drink from various Pacific islands with anxiolytic properties. Multiple meta-analyses including a Cochrane review find it to be an effective anxiety treatment, but its safety is in question after reports of several cases of liver failure caused by the plant. This may be yet another case of people exaggerating freakishly rare side effects; the risk has been estimated at less than one in a million doses (though remember that if you take it daily for ten years, that number bcomes 1/300). Others suggest a rate as low as one in a hundred million but this assumes zero underreporting; others challenge this assumption. Possibly it is only poorly prepared kava causes liver problems; for traditionally prepared kava, look for preparations that specify they are made from root/rhizome material only. The American Academy of Family Physicians recommends that:

Physicians who supervise patients taking kava for the treatment of GAD should take care to avoid the following: (1) high dosages (more than 300 mg per day); (2) combining kava with hepatoactive agents; (3) using non-root preparations; and (4) exposure for longer than 24 weeks. Use of WS1490 standardized kava extract is also recommended. If these safety precautions are followed, kava can be appropriate therapy for selected patients diagnosed with GAD

Don’t take kava if you have any liver problems, if you’re on any medications that might interact with it, or if you plan on drinking alcohol at the same time. Consider talking about it with your doctor first and getting plans to check liver enzymes regularly.

Selank is an experimental Russian anti-anxiety medication going through their version of clinical trials. It’s a bit high-maintenance – you have to keep it refrigerated or else it decays, and the only two functional means of administration are injection or nasal spray – but anecdotal evidence is extraordinarily positive. No side effects have been found thus far, but needless to say by the time you get to “injecting experimental Russian medications into yourself” we have left the point where we can entirely guarantee this is a good idea. Ceretropic sells a nasal spray version, which is probably more convenient than having to inject it.

Phenibut is another Russian anti-anxiety medication. It is potentially addictive and dangerous. I do not want to actively recommend against it, because it can be very useful if used infrequently and carefully. Discussing exactly how to use it infrequently and carefully is beyond the scope of this article. Please do not use this unless you have looked into it carefully and understand the risks and benefits.

Overall, the best evidence seems to be for l-theanine (especially if you drink coffee) and bacopa (especially if you’re willing to wait months for any effect), with picamilon also worth your time to try and Selank as an option for the very adventurous.

V. Conclusions

No treatment stands out as extremely effective, and the best route to dealing with anxiety probably depends on many factors like your amount of free time, your motivation, your access to medical care, and your willingness to put up with side effects. After you’ve fixed lifestyle issues, I think any of “self-help workbook”, “start SSRIs”, or “try l-theanine” are good first options. On the other hand, benzodiazepines, antipsychotics, and kava are all options I would hold off on until you’ve tried a couple of other things.

Like with the depression post, the most important conclusion you can take from this is that you have lots of options. Please don’t let people give you an SSRI and then give up. Work with your doctor. Anxiety actually has a pretty good prognosis if people work on it, but it can be a difficult and frustrating process. Just remember: there are lots of options.

PS: Relevant Onion

# CBT In The Water Supply

[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 of 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 ot 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 as only 20%. This was surprising, since he’d received a thirty-second standing obation, 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, the 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 story 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. How 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 saidL “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 wehat 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 your’e 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 fromj 80% to 20%. His felings of guilt, shame, and anxiety fell all the way to 10%, and his feelings of hopelessness dropped to 5%. The feelings of loneliness, embarassment, 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:

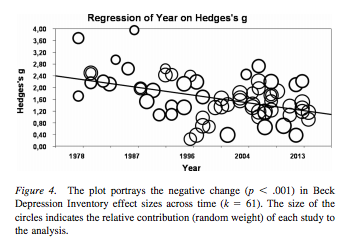
[When I first learned about this therapy, I thought] depression and anxiety seemed far too serious and severe for such a simplistic approach. But when I tried these methods with some of my more difficult patients, my perceptions changed. Patients who’d felt hopeless, worthless, and desperate began to 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 as 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.



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 has 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]

# Reverse Psychology

[Content warning: suicide]

I.

It all started when I made that phone call.

I was really bad. All the tenure-track positions I’d applied to had politely declined, and I saw my future in academia gradually slipping away from me. Then the night before, my boyfriend had said he thought maybe we should start seeing other people. I didn’t even know if we were broken up or not, and at that point I couldn’t bring myself to care. I sat on my bed, thinking about things for a while, and finally I called the suicide hotline.

“Hello?” a woman’s voice answered on the other side. Somehow, just hearing someone else made me feel about five times better.

“Hello,” I said, a little more confidently. “I’ve been thinking of committing suicide. I need help.”

“Okay,” she said. “Is there a gun in your house?”

“No.”

“All right. The first thing you need to do is get one. Overdosing on pills is common, but it almost never works. You can get a firearm at almost any large sporting goods store, but if there aren’t any near you, we can start talking about maybe jumping from a high…”

“What the HELL?” I interrupted, suddenly way more angry than depressed. “You’re supposed to @#!$ing tell me not to do it!”

“This is the suicide hotline,” the woman said, now sounding confused. Then, “Are you sure you weren’t thinking of the suicide prevention hotline?”

“Give me a break! I took a psychology class in undergrad, I know what a suicide hotline is!”

“I’m sorry you seem to be upset. But this is the suicide hotline. It’s like how there’s the Walk For Breast Cancer, but also the Walk Against Breast Cancer.”

“There’s the what? But…I was in the Walk For Breast Cancer! I thought…”

“It sounds like you have some issues,” said the woman, politely.

“Ugh,” I said. “Yeah.”

“Do you feel like you need professional help?”

“Yeah.”

“I do have a free clinic with an opening available tomorrow at three PM, would you like me to slot you in for an appointment?”

So you’re probably wondering why in the world I would take an appointment arranged by the suicide hotline that wasn’t a suicide prevention hotline. The answer is – were you even listening? A free clinic? With an appointment available the next day? Normally I was lucky if I found a place with an opening in less than two months and a co-pay that wasn’t completely ruinious. You bet I was taking that appointment before someone else snatched it up.

Dr. Trauer’s office looked gratifyingly normal. There was a houseplant, a diagram of the cranial nerves, some Abilify® merchandise, and on the wall one of those Magic Eye stereographic images that resolved into a 3D picture of the human brain. Dr. Trauer himself looked like your average doctor – a little past middle age, a little overweight, a short greying beard. He motioned me to sit down and took the paperwork I’d been filling out.

“Hmmmm,” he said, reading it over. “29 years old, postdoc in biochem, recent relationship trouble…mmmm…you did the right thing.”

“In coming here?”

“No, in considering suicide. After getting rejected from a tenure-track position, your life is pretty much over.”

“WHAT?”

“I mean, here you are, hundreds of thousands of dollars in debt, with only one area of expertise, and now you’ve been rejected from it. I can totally see why you might think it’s worth ending it all.”

“But…there are lots of other things I can do! I can get a job in industry! I can work in something else! Even if I can’t find a job right away, I have parents who can help support me.”

“Industry!” Dr. Trauer was having none of it. “A bunch of bloodsuckers. Do you realize how bad work in the private sector is these days? They’ll abuse you and then spit you out, and once you’ve been out of university too long nobody else will want you.”

“Lots of people want biochemists! If I work for a company for a few years, I’ll have more experience and maybe that will make me more attractive to employers! What…what kind of a psychiatrist are you, anyway?”

“Cindy didn’t tell you?”

“Cindy?”

“The woman on the phone.”

“She didn’t really tell me anything!”

“Well,” said Dr. Trauer. “To answer your question, we’re dark side psychiatrists. This is the state’s only dark side psychiatry clinic.”

“Dark side psychiatry? Really?”

“We’re a…well, some people say sect, but I like to think of it as more of a guild…dedicated to improving negative mental health. Think of it this way. When you’re a hijacked murder-monkey hurtling toward your inevitable death, sanity is a completely ridiculous thing to have. And when the universe is fifteen billion light-years across and almost entirely freezing void, the idea that people should have ‘coping skills’ boggles the imagination. An emotionally healthy person is a person who isn’t paying attention, and our job is to cure them.”

“There’s more than one of you?”

“Oh, yes. There’s a thriving dark side psychiatric community. There are dark side psychopharmacologists – you’d be amazed what a few doses of datura can do to a person. There are dark side psychotherapists who analyze and break down people’s positive cognitions. There are dark side child psychiatrists who catch people when they’re young, before sanity has had a chance to take root and worsen. And there are dark side geriatric psychiatrists, who go from nursing home to nursing home, making sure that the elderly are not warehoused and neglected at exactly the time it is most important to ensure that stroke or dementia does not protect them from acute awareness of the nearness of death.”

“That’s awful!” I said.

“Is it? Look where sanity’s gotten you. You want to kill yourself, but you don’t have the courage. Work with me for ten sessions, and I promise you we can help you get that courage.”

“You’re a @#!$ing quack,” I said. “And if you think killing yourself is so great, how come you haven’t done it yourself yet?”

“Who says I haven’t?” asked Dr. Trauer.

His hand went to his face, and he plucked out his right eye, revealing an empty void surrounded by the bleached whiteness of bone. I screamed and ran out of the clinic and didn’t stop running until I was in my house and had locked the door beside me.

II.

“…and that’s pretty much the whole story, doctor,” she told me. “And then I looked to see if there were any real psychiatrists in the area and someone referred me to you.”

“Well,” I said, my face unreadable. “I can certainly see why you’re complaining of, how did you put it, ‘depression and acute stress disorder’.”

“Not so acute anymore. It took me two months to get an appointment at your clinic.”

“Oh,” I said. Then, “Sorry, we’re sort of backed up.” Then, “Okay. We’ve got a lot we have to work on here. Let me tell you how we’re going to do it. We’re going to use a form of therapy that challenges your negative cognitions. We’re going to take the things that are bothering you, examine the evidence for them, and see if there are alternative explanations.”

“What do you mean?” she asked.

“Well,” I said. “It seems to be this Dr. Trauer incident that’s traumatized you a lot. I can see why you would be stressed out. The way you tell it, it sounds absolutely terrifying.”

“You don’t believe me,” she said, not accusatory, just stating a fact.

“I think it would be helpful to examine alternate explanations,” I said. “I’m willing to assume it happened exactly as you tell it. I can see why you would think Dr. Trauer wanted you to commit suicide. But are there any alternative explanations for the same event?”

“I don’t see how there can be,” she said. “He outright said that he thought I should kill myself.”

“Right. But from what you know of psychiatrists and therapy – and you did say you took some classes in undergrad – are there any other reasons he might have said something like that?”

She thought for a second. “Wait,” she told me. “There’s a technique in therapy called paradoxical intention. Where you take a patient’s irrational thought, and then defend and amplify it. And then when the patient hears it from someone else, she realizes how silly it sounds and starts arguing against it, and then it’s really hard to keep believing it after you’ve shot it down yourself.”

I nodded. “That’s definitely a therapeutic method, and sometimes a very effective one. Do you have any evidence that this is what Dr. Trauer was doing?”

“Yes! As soon as he said I should commit suicide, I started arguing against him. He told me that if I couldn’t get a tenure track position there would be no other jobs available, and I told him there would be! Then he told me that the jobs would be terrible and I’d never be able to make a happy life for myself with them, and I argued that I would! That must have been what he was going for!”

She suddenly looked really excited. Then, just as suddenly, the worry returned to her face.

“But then what happened with his eye? I swear I saw him take it right out of the socket.”

I nodded. “Can you think of any alternate explanations for that?”

Thinking about it that way, it only took her like five seconds. She slapped her head like she’d been an idiot. “A glass eye. He probably had some kind of injury, had to put in a glass eye, and could take it out any time he wanted. He must have thought it would be a funny gag and didn’t realize how traumatized I’d be. Or he wanted to scare me into realizing how much I wanted to live. Or something.”

I nodded. “That does sound like a reasonable explanation.”

“But…don’t people with glass eyes usually have like scar tissue and normal skin behind them? This guy, I swear it was just the bone and this empty socket, like you were seeing straight to his skull.”

“You’re asking the right questions,” I said. “Now think a little more.”

“Hmmmm,” she said. “I guess I was really, really stressed out at the time. And I only saw it for, like, a fraction of a second. Maybe my brain was playing tricks on me.”

“That can definitely happen,” I agreed.

She looked a lot better now. “I owe you a lot of thanks,” she said. “I’ve only been here for, like, fifteen minutes, and already I think a lot of my stress has gone away. All of this really makes sense. That paradoxical intention thing is actually kind of brilliant. And I can’t deny that it worked – I haven’t been suicidal since I talked to the guy. In fact…okay, this is going to sound really strange, but…maybe I should go back to Dr. Trauer.”

I wrinkled my forehead.

“It’s not that I don’t like you,” she said. “But he had this amazing free clinic, and what he did for me that day…now that I realize what was going on, that was actually pretty incredible.”

“Hold on a second,” I said.

I left the room, marched up to the front desk, took the directory of medical providers in the area off the shelf, marched back to the room. I started flipping through the pages. It was in alphabetical order…Tang…Thompson…Tophet…there we go. Trauer. My gaze lingered there maybe just a second too long, and she asked if I was okay.

“Um, yeah,” I said. “It’s just that he doesn’t – he doesn’t take your insurance. That’s the problem.”

“It’s okay,” she told me. “He said it was a free clinic. So that shouldn’t a problem.”

“Well, uh…the thing is…when you see out-of-network providers, your insurance actually charges, charges an extra fee. Even if the visit itself is free.”

She looked skeptical. “I’ve never heard of that.”

“It’s new. With Obamacare.”

“Really? How high a fee is it?”

“It’s…um…ten thousand dollars. Yeah, I know, right? Thanks, Obama.”

“Wow,” she said. “I definitely can’t afford that. I guess I’ll keep coming here. Not that there’s anything wrong with that. You’ve been very nice. It’s just that…with Dr. Trauer…well…sorry, I’ll stop talking now. Thanks a lot, doctor.” She stood up and shook my hand before heading for the door. “Seriously, I can’t believe how much you’ve helped me.”

No, I thought, as she departed you can’t. I told her she was asking the right questions, and she was, but not all of them.

For example, why would a man with only one working eye have a stereographic Magic Eye image in his office?

I picked up my provider directory again, stared a second time at the entry for Dr. Trauer. There was a neat line through it in red pen, and above, in my secretary’s careful handwriting, “DECEASED”.

Before returning the directory to the front desk, I took my own pen and added “DO NOT REFER” in big letters underneath.

# Freedom On The Centralized Web

I.

A lot of libertarians and anarcho-capitalists envision a future of small corporate states competing for migrants and capital by trying to have the best policies.

But the Internet is about as close to that vision as we’re likely to find outside the pages of a political philosophy textbook. And I am far from convinced.

Let’s back up. Internet communities – ranging from a personal blog like this one all the way up to Facebook and Reddit – share many features with real communities. They work out rules for punishing defectors – your trolls, your harassers – and appoint a hierarchy of trusted individuals to carry out those rules. They try to balance competing concerns like free expression and public decency. They host cliques, power grabs, flame wars, even religious strife. They try to raise revenue, they establish a class system of Power Users and Premium Users, they deal with resentment from people who aren’t getting their way. They develop a culture.

The job of a community leader, be they a blogger or the CEO of Facebook, is a lot like the job of the Mayor of New York City: create a pleasant community where talented people will want to live and work, where wrongdoing is met with swift punishment, and where you can collect revenue without annoying your constitutents too much. But it’s even more like a hypothetical corporate state CEO in a Patchwork or Archipelago – wield absolute power, tempered by the knowledge that your citizens can leave at any time – and if they don’t, skim a little off the top of their productive activity.

In theory, this is supposed to lead to amazing communities as corporate states optimize themselves to get more customer-citizens and new polities arise to take advantage of deficiencies in the old.

In practice, we tried this with the Internet for a couple of years, and then moved to the current system, where individual sites like blogs and little storefronts are in decline and conversation and commerce have moved to a couple of giant corporations: Facebook, Twitter, Reddit, Amazon, Paypal.

These companies aren’t exactly monopolies. To some degree, if you’re unsatisfied with Facebook you can move to Twitter. But they’re not exactly competitors either – there are a lot of things Facebook is good for that Twitter fails completely, and vice versa. It’s like Coca-Cola vs. milk: in theory you’ve always got the choice to drink either in place of the other; in practice you usually know which one you need at any given time. In that sense, there’s no real Facebook competitor except eg Orkut or Diaspora, which no one uses.

Which suggests one reason why these sites are so dominant: their main selling point is their size. Facebook is the best because all of your friends are on it; if I made a much better Facebook clone tomorrow no one would go unless everyone else was already there (Google found this out the hard way). Amazon is the best because you can buy pretty much everything you want there; Paypal is the best because most sites take PayPal. So not only do they have no competitors, but it’s really hard to imagine one ever arising. In order to compete with Facebook, you not only need a better product, you need a product that’s so much better that everybody decides to switch en masse at the same time. The only example I can think of where this ever worked was the Great Digg Exodus, where Digg screwed up their product so thoroughly that everyone simultaneously said “@#!$ this” and moved to Reddit.

So instead of “let a thousand nations bloom”, it ended up more like “let five or six big nations bloom that we can never get rid of”.

II.

It’s a truism that the First Amendment only protects citizens from the government, not from other citizens. Nothing stops a private college from expelling any student who criticizes the administration, and nothing stops a private business from firing any employee who doesn’t support the boss’ preferred candidate. We apparently place our trust in the multiplicity of the market to maintain some semblance of freedom; out of thousands of competing companies, not all will ban the same political positions; if too many did so, other companies would start offering freedom of speech as a benefit and poach the more repressive companies’ employees and customers.

It’s a little concerning that we accept this argument about freedom of speech when we don’t accept it for anything else. We don’t trust the free market to necessarily preserve racial equality – that’s what anti-discrimination laws are for. We don’t trust the free market to necessarily preserve worker safety – that’s what OSHA and related regulations are for. We don’t even trust the free market to necessarily preserve fire safety – that’s why federal inspectors have to come in every so often to make sure you’re not secretly plotting to let your employees fry. Whenever we think something is important, we regulate the hell out of it, rights-of-private-companies to-set-their-own-policies be damned. But free speech? If you don’t trust the free market to sort it out, the only possible explanation is that you just don’t understand the literal text of the First Amendment.

The argument for non-discrimination laws is that discrimination isn’t just random noise. If a couple of companies here and there decided to discriminate, then they might be easily overtaken by nimbler companies willing to take any employees and customers who came to them; and even if they didn’t, a couple of companies here and there discriminating wouldn’t be the end of the world. The argument for non-discrimination laws is that discrimination can take the form of global social pressure in favor of discrimination, enforced by punishing defectors, to the point where certain races can find themselves locked out of the economy altogether.

Concerns about freedom of speech come from much the same place. Back when homosexuality was really taboo, you’d have a very tough time finding any reference to it, let alone a positive reference to it, in any newspaper or TV channel in the country. All the big companies knew that talking about it (or letting their editorial staff talk about it) was the sort of thing that could get them in trouble, and they had no particular incentive to do so – so they didn’t. Yes, eventually they reversed that policy, but I’m not exactly going to be able to cite an example that didn’t later become okay and still have everyone believe it’s a good example of something it was wrong to have banned!

But even when homosexuality was banned from formal discussion on the news, there was still the opportunity to discuss it with your friends in private. I don’t know much about the history of the gay rights movement, but I understand it was a few small groups of like-minded people who managed to coordinate such discussions among themselves using non-mass-media that started some of the activism that eventually led to it become accepted more generally.

Nowadays that’s a little more complicated. If every company in the world decided that their profit margin required them to appear Tough On Homosexuality, it wouldn’t just mean no mass media editorials. Insofar as a lot of the public square has been annexed by Facebook and Twitter and Reddit, the discussion can be kept out of the public square in a way it couldn’t have been previously. Insofar as the economy relies on PayPal and Amazon as a currency system and marketplace respectively, companies can just decide that currency cannot be used to support gay rights, in much the same way that for a while currency could not be used to support WikiLeaks. The nuclear option is that Google decides not to show gay-related sites in its search results, so that you could make as many persuasive arguments for legalizing homosexuality as you want and no one would ever find them unless you knock on their door and hand them the URL directly.

(The thermonuclear option is that browsers just include some code to refuse to render any site relating to homosexuality, and now you’re done. But that is ridiculous – who would ever believe that browser companies would take it upon themselves to be the arbiter of people’s personal beliefs about homosexuality?)

This is not entirely theoretical. You want some really weird porn? You probably won’t find it on Amazon, according to the delightfully-named article Amazon’s War On Bigfoot Erotica. After they got bad press for hosting some kind of out-there stuff, they decided that anything which offended too many people’s sensibilities was a liability. This echoes a much more serious decision from a few years earlier: Paypal threatened to suspend the accounts of any companies selling sufficiently gross erotic books. Booksellers, many of whom made only a tiny percent of their profit from erotica, claimed that their hands were tied; if you can’t use PayPal, selling on the Internet suddenly becomes a much more dubious proposition. This story has a happy ending; Paypal eventually amended their policy to limit it to much more specific cases. But for a while, it was touch-and-go enough that a few people started wondering: “Hey, maybe we shouldn’t have entrusted our entire commercial infrastructure to a private company with no accountability.”

Advocates of net neutrality like to worry about a “two-tiered” Internet, where the companies that can make sweetheart deals with the ISPs are easy for everyone to access, and everybody else can only be accessed with a bit more money and a bit more trouble. Well, I worry about a two-tiered marketplace of ideas. Write decent erotica, socially approved erotica where everyone has heterosexual sex and then goes to church afterwards, and you can sell it on Amazon, collect profits using PayPal, talk to your friends about it on Facebook, and advertise on Reddit. Write weird erotica, the kind that other people might find offensive, and you might have to start your own website, take payment via some inconvenient method like Bitcoin, have trouble advertising it by word of mouth, and not be able to talk about it on literary discussion forums. It’s not that you’ve been banned from writing your erotica. You can write it. It’s just that practically nobody else will ever hear about it or buy it, except maybe the tiny fraction of people who are already extremely clued-in to the weird erotica scene and know exactly where to look for it.

This isn’t so much different from the old days when nobody would talk about homosexuality. Indeed, one could argue that the modern world is friendlier to people with unpopular ideas – there are more opportunities to self-publish, to bypass traditional bookstores, and to get covered in weird niche news outlets.

But at the same time, the amount of the information ecology controlled by private companies has increased drastically, and if private companies don’t like you, now you have entirely new problems.

III.

I used to think that there was enough demand for a free marketplace of ideas that if a company become too restrictive, another one would spring up to replace it. Then I suffered through the conflict between Reddit and Voat.

Reddit recently alienated (no pun intended) some of its users, who decided to move en masse to an alternative Reddit-like platform called Voat, whose owner promised not to restrict content unless it was illegal (in his home country of Switzerland, which permits a lot). I don’t want to get into the details too much (though I did explain my perspective on it on Tumblr), but suffice it to say that (one) (small) part of the problem was that people thought Reddit was failing its free speech principles by cracking down on various unsavory groups.

HL Mencken once said that “the trouble with fighting for human freedom is that one spends most of one’s time defending scoundrels. For it is against scoundrels that oppressive laws are first aimed, and oppression must be stopped at the beginning if it is to be stopped at all.”

There’s an unfortunate corollary to this, which is that if you try to create a libertarian paradise, you will attract three deeply virtuous people with a strong committment to the principle of universal freedom, plus millions of scoundrels. Declare that you’re going to stop holding witch hunts, and your coalition is certain to include more than its share of witches.

So while some small percent of Reddit’s average users moved over, a very large percent of its witches did. Sometimes the witchcraft was nothing worse than questioning Reddit’s political consensus. Other times, it was harassment, hate groups, and creepy porn.

(I don’t want to get into the eternal “you’re hosting child porn!” versus “photos of clothed fifteen year olds aren’t child porn, they’re perfectly fine!” debate, except to say that when the universe finally runs down, and we all succumb to entropy, the second-to-last post on the ultra-cyber-quantum-internet will be “posting holograms of neotenous transhumans is totally in conformity with the First Law Of Robotics as long as they are older than thirteen million years and created the hologram themselves”, and the last post will be “lol u r a perv”)

I feel obligated to say that, in spite of CONSTANT MEDIA SMEARS, Reddit’s community is amazing, puts in astounding effort to help its members and fight for good causes all over the world, and that the representation of weirdoes and neotenous-transhuman-hologram people is no higher than any other part of the population. But that’s not zero. And a disproportionate number of those people became interested in the new site.

Already, we see why the typical answer “If you don’t like your community, just leave and start a new one” is an oversimplification. A community run on Voat’s rules with Reddit userbase would probably be a pretty nice place. A community run on Voat’s rules with the subsection of Reddit’s userbase who will leave Reddit when you create it is…a very different community. Remember that whole post on Moloch? Even if everyone on Reddit agrees in preferring Voat to Reddit, it might be impossible to implement the move, because unless everybody can coordinate it’s always going to be the witches who move over first, and nobody wants to move to a community that’s mostly-witch.

But the problem isn’t just natural self-sorting. The problem is natural self-sorting, plus enemy action. Remember, the big corporations do what they do because it’s what everyone in society is demanding. To break from that mold is to pretty much set yourself up as everyone’s enemy and invite retaliation. The media and Reddit’s SJ community quickly denounced Voat as Public Enemy No 1; as a result, in its first week it got DDoS attacked, deleted by its hosting company with no explanation except “the content on your server includes politically incorrect parts”, and had its PayPal account frozen. As a result, the Great Reddit Exodus was placed on hold while they tried to get their site back up, and by the time they did Reddit had switched CEOs and the momentum was gone.

Advocates of free-market governance and “let a thousand nations bloom” like to talk as if overly restrictive laws in one polity will immediately result in the rise of other competing policies that throw off their shackles and outcompete the first. But even on the relatively lawless Internet, where startup costs are so low that a random student from Switzerland can decide on a whim to take on one of the largest websites in the world, it’s way more complicated than that.

IV.

Actually, the whole Reddit thing left a bad taste in my mouth.

It would be paranoid to say that there are people for whom fighting against free speech is a terminal value, but let me make a slightly weaker claim. There are people who consider themselves the protectors of decency, who notice that their opponents are usually using the value “free speech” to oppose their demands, and so “free speech” to these people becomes the equivalent of “small government” or “tolerance and equality” or “family values” – a value which most people agree is good, but which has gotten claimed by one side of a political argument so hard that for the other side it becomes an outgroup signal and sign of cringeworthy bad arguments which must be shot down. These people don’t quite have fighting free speech as a terminal value, but you might as well model them as if they do. These are the people who say “freeze peach” in the same way other people say “but mah jawbs!”

And these people have a winning strategy. I’ve seen it with Reddit and any other website that gets on their bad side. The strategy is weaponized stereotype campaigns. If a site tolerates witches, describe it as a witch site about witchcraft populated entirely by witches. It’s super easy. By happy coincidence, Slate even has an article calling people out on it this very week.

Think about it like this. No matter how many brilliant artists, scientists, and humanitarians Islam produces, in the mind of a good chunk of Westerners it will always be associated first and foremost with terrorism. Redditors, Diggians, Tumblrites, 4chanistas, Instagramastanis, Slashdotmen, Metafilterniks – all are groups that the average person knows a whole lot less about than they do Muslims. A concerted campaign to irrevocably identify an entire online community with a few atrocious actions by its worst members will succeed pretty much instantly. There are 36 million Redditors, so unless they advertise solely in the saint demographic, we expect the worst members to be pretty bad. Therefore, Reddit is at the mercy of anyone with the resources to start such a campaign. Reddit Inc’s main asset is its brand, so it has every incentive to cave – even a principled leadership would rather make a few administrative changes than sacrifice the whole to save some Holocaust deniers or whatever.

After that, the site’s userbase has two options – either suck it up, or go off somewhere else. Go off somewhere else, and they’ll get DDoSed, taken down by their host, and slowly starved of money like Voat, at the same time as the same media forces accuse the new site of being a hot spot for witchcraft – this time with good reason. The new site might not die out completely, but it will be sufficiently established in the hearts of everyone as a Bad Place that it will be stuck in the same equilibrium as central Detroit – only people with no other options will go there, because it is inhabited mostly by the sort of people with no other options.

The worst possible end-game for this is the two-tier marketplace of ideas mentioned above, with an unfortunate twist – everyone knows that the second tier is inhabited entirely by witches, and therefore being on the second tier is sufficient to convict you. Unpopular ideas are gradually forced out of the first tier by media smear campaigns, and from then on everyone believes the effort was justified, because it’s one of those second-tier ideas that you only find in the same sites as the racists and trolls and child pornographers. You’re not a second tier kind of person, are you? No, we didn’t think so.

I have no particular solution to this. Certainly the well-intentioned solutions other people are working on, like a decentralized crypto-Reddit that can’t be moderated even in principle, are unlikely to help (hint: what is the most striking difference between Bitcoin marketplaces and normal marketplaces?) My primary hope is that it’s just not a real problem. Certainly there has been very little in the way of speech restriction so far, and what little there has been has been against things which, on the object level, I’m happy to see gone. It’s entirely possible that we’ll escape with only a few things banned that probably deserve it. I certainly hope this is the case.

I’m just annoyed that we’ve gotten ourselves in a corner where we have to depend on hope.

# The General Factor Of Correctness

People on Tumblr are discussing Eliezer Yudkowsky’s old essay The Correct Contrarian Cluster, and my interpretation was different enough that I thought it might be worth spelling out. So here it is: is there a General Factor of Correctness?

Remember, IQ is supposed to come from a General Factor Of Intelligence. If you make people take a lot of different tests of a lot of different types, people who do well on one type will do well on other types more often than chance. You can do this with other things too, like make a General Factor Of Social Development. If you’re really cool, you can even correlate the General Factor of Intelligence and the General Factor of Social Development together.

A General Factor Of Correctness would mean that if you asked people’s opinions on a bunch of controversial questions, like “Would increasing the minimum wage to $15 worsen unemployment?” or “Which interpretation of quantum mechanics is correct?” or “are artificial sweeteners safe?” and then somehow discovered the answers to these questions, people who did well on one such question would do well on other types more often than chance.

This is a surprisingly deep and controversial issue, but one with potentially big payoffs. Suppose you want to know whose economic theories are right, but you don’t want to take the time to learn economics. Consider some position that was once considered fringe and bizarre, but now known to be likely true – for example, pre-Clovis settlement of the New World. Find the economists who believed in pre-Clovis settlement of the New World back when doing so was unpopular. Those economists have demonstrated a proven track record of being able to winnow out correct ideas amidst a sea of uncertainty. Invest in whatever company they tell you to invest in and make a killing.

I’m sort of joking, but also sort of serious – shouldn’t something like this work? If there’s such a thing as reasoning ability, people who are good at sifting through a mess of competing claims about pre-Columbian anthropology and turning up the truth should be able to apply that same skill to sifting through a mess of competing claims about economic data. Right?

If this is true, we can gain new insight into all of our conundra just by seeing who believes what about New World migration. That sounds useful. The problem is, to identify it we have to separate it out from a lot of closely related concepts.

The first problem: if you just mark who’s right and wrong about each controversial issue, the General Factor Of Correctness will end up looking a lot like a General Factor of Agreeing With Expert Consensus. The current best-known heuristic is “always agree with expert consensus on everything”; people who follow this heuristic all the time are most likely to do well, but we learn nothing whatsoever from their success. If I can get brilliant-economist-points for saying things like “black holes exist” or “9-11 was not a government conspiracy”, then that just makes a mockery of the whole system. Indeed, our whole point in this exercise is to see if we can improve on the “agree with experts” heuristic.

We could get more interesting results by analyzing only people’s deviations from expert consensus. If you agree with the consensus about everything, you don’t get to play. If you disagree with the consensus about some things, then you get positive points when you’re right and negative points when you’re wrong. If someone ends consistently ends up with a positive score beyond what we would expect by chance, then they’re the equivalent of the economist who was surprisingly prescient about pre-Clovis migration – a person who’s demonstrating a special ability that allows them to outperform experts. This is why Eliezer very reasonably talks about a correct contrarian cluster instead of a correct cluster in general. We already know who the correct cluster is, and all of you saying “I have no idea what Clovis is, but whatever leading anthropologists think, I think that too” are in it. So what? So nothing.

The second problem: are you just going to rediscover some factor we already know about, like IQ or general-well-educatedness? I’m not sure. WHen I brought this up on Tumblr, people were quick to point out examples of very intelligent, very well-educated people believing stupid things – for example, Newton’s obsession with alchemy and Biblical prophecy, or Linus Pauling’s belief that you could solve health just be making everyone take crazy amounts of Vitamin C. These points are well-taken, but I can’t help wondering if there’s selection bias in bringing them up. Yes, some smart people believe stupid things, but maybe even more stupid people do? By analogy, many people who are brilliant at math are terrible at language, and we can all think of salient examples, but psychometrics has shown again and again that in general math and language skills are correlated.

If we look for more general data, we get inconsistent results. Neither IQ nor educational attainment seems to affect whether you believe in climate change very much, though you can get slightly different results depending on how you ask and what you adjust for. There seems to be a stronger effect of intelligence increasing comfort with nuclear power. Other polls show IQ may increase atheism, non-racism, and a complicated cluster of political views possibly corresponding to libertarianism but also showing up as “liberalism” or “conservativism” depending on how you define your constructs and which aspects of politics you focus on. I am very suspicious about any of this reflecting real improved decision-making capacity as opposed to just attempts to signal intelligence in various ways.

The third problem: can we differentiate positive from negative selection? There are lots of people who believe in Bigfoot and ESP and astrology. I suspect these people will be worse at other things, including predicting economic trends, predicting world events, and being on the right side of difficult scientific controversies, probably in a way independent of IQ or education. I’m not sure of this. But I suspect it. If I’m right, then the data will show a General Factor of Correctness, but it won’t necessarily be a very interesting one. To give a reductio ad absurdum, if you have some mental disorder that causes you to live in a completely delusional fantasy world, you will have incorrect opinions about everything at once, which looks highly correlated, but this doesn’t necessarily prove that there are correlations among the people who are more correct than average.

The fourth problem: is there a difference between correctness and probability calibration? Suppose that Alice says that there’s a 90% chance the Greek economy will implode, and Bob has the same information but says there’s only an 80% chance. Here it might be tempting to say that one of either Alice or Bob is miscalibrated – either Alice is overconfident or Bob is underconfident. But suppose Alice says that there’s a 90% chance the Greek economy will implode, and Bob has the same information but says there’s only a 10% chance that it will. Now we’re more likely to interpret this in terms of them just disagreeing. But I don’t know enough about probability theory to put my finger on whether there’s a true qualitative difference.

This is important because we know calibration is a real thing and some people are good at it and other people aren’t but can improve with practice. If all we’re showing is that people who are good with probabilities are good with probabilities, then whatever.

But there are tantalizing signs that there might be something more here. I was involved in an unpublished study which I can’t upload because I don’t have the other authors’ permission, but which showed conclusively that people with poor calibration are more likely to believe in the paranormal (p < 0.001), even when belief in the paranormal was not assessed as a calibration question.  
  
So I went through the Less Wrong Survey data, made up a very ad hoc measure of total calibration skill, and checked to see what it did and didn't predict. Calibration was correlated with IQ (0.14, p = 0.01). But it was also correlated with higher belief in global warming (0.13, p = 0.01), with higher belief in near-term global catastrophic risk (-0.08, p - 0.01), increased support for immigration (0.06, p = 0.048) and with decreased support for the human biodiversity movement (0.1, p = 0.002). These were all independent of the IQ correlation. Notably, although warming and GCR were asked in the form of probabilities, immigration and HBD weren't, suggesting that calibration can be (weakly) correlated with opinions on a non-calibration task.   
  
Maybe the most intriguing evidence for a full-fledged General Factor of Correctness comes from Philip Tetlock and IARPA's Good Judgment Project, which got a few thousand average people and asked them to predict the probability of important international events like “North Korea launches a new kind of missile.” They found that the same small group of people consistently outperformed everyone else in a way incompatible with chance. These people were not necessarily very well-educated and didn’t have much domain-specific knowledge in international relations – the one profiled on NPR was a pharmacist who said she “didn’t know a lot about international affairs [and] hadn’t taken much math in school” – but they were reportedly able to outperform professional CIA analysts armed with extra classified information by as much as 30%.

These people aren’t succeeding because they parrot the experts, they’re not succeeding because they have more IQ or education, and they’re not succeeding in some kind of trivial way like rejecting things that will never happen. Although the article doesn’t specify, I think they’re doing something more than just being well-calibrated. They seem to be succeeding through some mysterious quality totally separate from all of these things.

But only on questions about international affairs. What I’d love to see next is what happens when you ask these same people to predict sports games, industry trends, the mean global temperature in 2030, or what the next space probe will find. If they can beat the experts in those fields, then I start really wondering what their position on the tax rate is and who they’re going to vote for for President.

Why am I going so into depth about an LW post from five years ago? I think in a sense this is the center of the entire rationalist project. If ability to evaluate evidence and come to accurate conclusions across a broad range of fields relies on some skill other than brute-forcing it with domain knowledge and IQ, some skill that looks like “rationality” broadly defined, then cultivating that skill starts to look like a pretty good idea.

Enrico Fermi said he was fascinated by the question of extraterrestrial life because whether it existed or it didn’t, either way was astounding. Maybe a paradox, but the same paradox seems true of the General Factor of Correctness.

Outside the Laboratory is a post about why the negative proposition – no such General Factor – should be astounding:

“Outside the laboratory, scientists are no wiser than anyone else.” Sometimes this proverb is spoken by scientists, humbly, sadly, to remind themselves of their own fallibility. Sometimes this proverb is said for rather less praiseworthy reasons, to devalue unwanted expert advice. Is the proverb true? Probably not in an absolute sense. It seems much too pessimistic to say that scientists are literally no wiser than average, that there is literally zero correlation.

But the proverb does appear true to some degree, and I propose that we should be very disturbed by this fact. We should not sigh, and shake our heads sadly. Rather we should sit bolt upright in alarm. Why? Well, suppose that an apprentice shepherd is laboriously trained to count sheep, as they pass in and out of a fold. Thus the shepherd knows when all the sheep have left, and when all the sheep have returned. Then you give the shepherd a few apples, and say: “How many apples?” But the shepherd stares at you blankly, because they weren’t trained to count apples – just sheep. You would probably suspect that the shepherd didn’t understand counting very well.

If, outside of their specialist field, some particular scientist is just as susceptible as anyone else to wacky ideas, then they probably never did understand why the scientific rules work. Maybe they can parrot back a bit of Popperian falsificationism; but they don’t understand on a deep level, the algebraic level of probability theory, the causal level of cognition-as-machinery. They’ve been trained to behave a certain way in the laboratory, but they don’t like to be constrained by evidence; when they go home, they take off the lab coat and relax with some comfortable nonsense. And yes, that does make me wonder if I can trust that scientist’s opinions even in their own field – especially when it comes to any controversial issue, any open question, anything that isn’t already nailed down by massive evidence and social convention.

Maybe we can beat the proverb – be rational in our personal lives, not just our professional lives.

And Correct Contrarian Cluster is about why the positive proposition should be equally astounding. If it’s true, you can gain a small but nonzero amount of information about the best economic theories by seeing what their originators predicted about migration patterns in pre-Columbian America. And you can try grinding your Correctness stat to improve your ability to make decisions in every domain of knowledge simultaneously.

I find research into intelligence more interesting than research into other things because improvements in intelligence can be leveraged to produce improvements in everything else. Research into correctness is one of the rare other fields that shares this quality, and I’m glad there are people like Tetlock working on it.

Discussion questions (adapted from Tumblr):

1. Five Thirty Eight is down the night before an election, so you search for some other good sites that interpret the polls. You find two. Both seem to be by amateurs, but both are well-designed and professional-looking and talk intelligently about things like sampling bias and such. The first site says the Blue Party will win by 5%; the second site says the Green Party will win by 5%. You look up the authors of the two sites, and find that the guy who wrote the first is a Young Earth Creationist. Do you have any opinion on who is going to win the election?

2. On the bus one day, you sit next to a strange man who mumbles about how Bigfoot caused 9-11 and the Ark of the Covenant is buried underneath EPCOT Center. You dismiss him and never see him again. A year later, you see on TV that new evidence confirms Bigfoot caused 9-11. Should you head to Florida and start digging?

3. Schmoeism and Anti-Schmoeism are two complicated and mutually exclusive economic theories that you don’t understand at all, but you know the economics profession is split about 50-50 between them. In 2005, a survey finds that 66% of Schmoeist economists and 33% of anti-Schmoeist economists believe in pre-Clovis settlement of the New World (p = 0.01). In 2015, new archaeological finds convincingly establish that such settlement existed. How strongly (if at all) do you now favor one theory over the other?

4. As with 3, but instead of merely being the pre-Clovis settlement of America, the survey asked about ten controversial questions in archaeology, anthropology, and historical scholarship, and the Schmoeists did significantly better than the anti-Schmoeists on 9 of them.

# Non-Dual Awareness

Seen on Lauren’s Facebook: How Does Academia Resemble A Drug Gang?

Their answer is that both academia and drug gangs are marked by an endless supply of foot soldiers willing to work in terrible conditions for a small chance at living the good life. In drug gangs, the average street-corner dealer makes $3-something an hour; given that he’s got a high chance of being arrested or shot, why doesn’t he switch to McDonalds instead where the pay’s twice as good and the environment’s a lot safer? The article suggests one reason is because drug gangs offer the chance of eventually becoming a drug kingpin who is drowning in money.

(I’d worry they’re exaggerating the importance of this factor compared to wanting to maintain street cred and McDonalds jobs being much more regimented both in the application process and performance, but they’re the ones who have talked to anthropologists embedded in drug gangs, not me.)

Academia has the same structure. TAs and grad students work in unpleasant conditions for much less than they could make in industry, because there’s always the chance they could become a tenured professor who gets to live the life of the mind and travel to conferences in far-off countries and get summer vacations off.

The article describes this structure as “dualization” – a field that separates neatly into a binary classification of winners and losers.

This concept applies much more broadly than just drugs and colleges. I sometimes compare my own career path, medicine, to that of my friends in computer programming. Medicine is very clearly dual – of the millions of pre-med students, some become doctors and at that moment have an almost-guaranteed good career, others can’t make it to that MD and have no relevant whatsoever in the industry. Computer science is very clearly non-dual; if you’re a crappy programmer, you’ll get a crappy job at a crappy company; if you’re a slightly better programmer, you’ll get a slightly better job at a slightly better company; if you’re a great programmer, you’ll get a great job at a great company (ideally). There’s no single bottleneck in computer programming where if you pass you’re set for life but if you fail you might as well find some other career path.

My first instinct is to think of non-dualized fields as healthy and dualized fields as messed up, for a couple of reasons.

First, in the dualized fields, you’re putting in a lot more risk. Sometimes this risk is handled well. For example, in medicine, most pre-med students don’t make it to doctor, but the bottleneck is early – acceptance to medical school. That means they fail fast and can start making alternate career plans. All they’ve lost is whatever time they put into taking pre-med classes in college. In Britain and Ireland, the system’s even better – you apply to med school right out of high school, so if you don’t get in you’ve got your whole college career to pivot to a focus on English or Engineering or whatever. But other fields handle this risk less well. For example, as I understand Law, you go to law school, and if all goes well a big firm offers to hire you around the time you graduate. If no big firm offers to hire you, your options are more limited. Problem is, you’ve sunk three years of your life and a lot of debt into learning that you’re not wanted. So the cost of dualization is littering the streets with the corpses of people who invested a lot of their resources into trying for the higher tier but never made it.

Second, dualized fields offer an inherent opportunity for oppression. We all know the stories of the adjunct professors shuttling between two or three colleges and barely making it on food stamps despite being very intelligent people who ought to be making it into high-paying industries. Likewise, medical residents can be worked 80 hour weeks, and I’ve heard that beginning lawyers have it little better. Because your entire career is concentrated on the hope of making it into the higher-tier, and the idea of not making it into the higher tier is too horrible to contemplate, and your superiors control whether you will make it into the higher tier or not, you will do whatever the heck your superiors say. A computer programmer who was asked to work 80 hour weeks could just say “thanks but no thanks” and find another company with saner policies.

(except in startups, but those bear a lot of the hallmarks of a dualized field with binary outcomes, including the promise of massive wealth for success)

Third, dualized fields are a lot more likely to become politicized. The limited high-tier positions are seen as spoils to be distributed, in contrast to the non-dual fields where good jobs are seen as opportunities to attract the most useful and skilled people. This reminds me of the other article I read today comparing academia to drug gangs, which was where Paul Krugman theorized that the reason so many criminals have horrible tattoos in inappropriate places is as a conspicuous symbol of criminality! He says that since these people’s tattoos mean they can never get a job in legitimate industry, other gang members and black market contacts can trust them to keep their bargains, since they’ve got no option under than continuing to work in the criminal underworld. Krugman writes (h/t Nathaniel Bechhofer):

The author, Diego Gambetta, adds a wonderful parallel: according to his account, Italian academics, who do a lot of horse-trading in appointments etc., cultivate a reputation for incompetence at actual research, again designed to reassure those with whom one deals.

“Being incompetent and displaying it,” he writes, “conveys the message I will not run away, for I have no strong legs to run anywhere else. In a corrupt academic market, being good at and interested in one’s own research, by contrast, signal a potential for a career independent of corrupt reciprocity. In the Italian academic world, the kakistrocrats are those who best assure others by displaying, through lack of competence and lack of interest in research, that they will comply with the pacts.”

(wait, this argument sounds kind of familiar. KRUGMAN, HAVE THEY GOTTEN TO YOU TOO?!)

II.

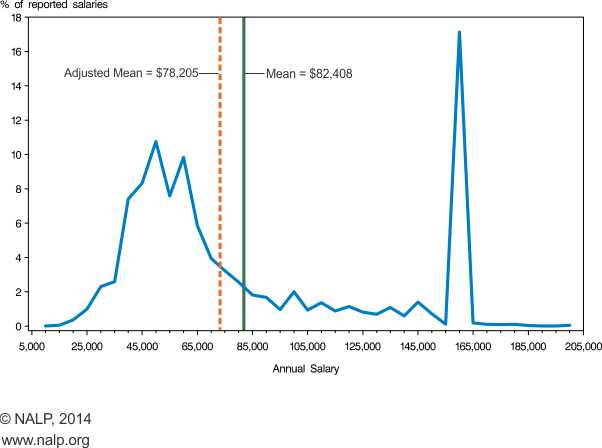
What dualizes some fields but not others?

Originally I was going to make a simplistic comment about licensing and regulation, but this doesn’t exactly capture it. Certainly the fact that medicine requires an MD has some effect on the dualization of medicine (alternatively, insofar as medicine isn’t entirely dualized, it’s because you can get less lucrative positions, like naturopath or therapist or nurse practitioner, without an MD). But we can imagine a system in which there were more than enough medical schools for everyone, anyone who applied to one got in, there was a glut of doctors, and the good doctors got good jobs and the less good doctors got less good jobs. So we might more soberly blame it on scarce licenses – for complicated reasons I won’t get into here, the number of residency spots is much lower than it should be, leading to a bottleneck where only a few people can obtain the MDs.

What about tenure? We can imagine an alternate universe where academia is populated with various PhDs on equal footing. Since there would be a glut, their salaries would be very low to start, but low salaries would mean easy employment, and colleges would find a lot of room for them to do one-on-one tutoring, or low-level research, or something like that. Eventually some of them would become a bit more prestigious in their fields and could demand higher salaries from hiring institutions, and a few superstars like Nobel Prize winners and the like could demand millions. At no point would there ever be anything called a “tenure track”. It seems like the main difference between this universe and our own is that tradition plus the reasonable desire of professors to be free from political interference has created this dichotomous variable called “tenure” and caused it to replace the continuous variable of salary as the prize for success. In favor of that theory, top professors seem weirdly underpaid compared to eg top athletes or top artists, even though I would expect having one of the world’s top scientists or historians to be a big draw for a school. According to the List Of Highest Paid Professors, only five professors in the US make more than a million dollars a year, and all of those are professors of lucrative medical subspecialties or of finance, who presumably are being paid that much to compensate them for teaching instead of participating in the high-paying professions they are otherwise qualified for.

What about drug dealers? I think there might be “licensing” at work here too. There’s no such thing as a mid-level independent drug dealer, because – if the three seasons I’ve watched of Breaking Bad are accurate – if you try this, the other drug dealers will shoot you. So you need a scarce “license” from the drug lords – basically El Chapo giving you the rights to a big piece of “turf” – instead of a license from the government. Whatever; I’m liberal Monday Wednesday Friday and libertarian Tuesdays and Thursdays; today is a Tuesday so all organizations that rely upon the use of force look the same to me.

But what about lawyers? Sure, there are regulations on who can practice, but the dualization in the legal fields comes after graduation of law school. Here, have some statistics:



The first step of what’s going on isn’t a mystery – the people on the very sharp mountain on the right are hired by big law firms on the “partner track” (note the similarity to “tenure-track”) and the people in the more gradual plateau on the left are everyone else. But there’s still a lot to be explained. Why isn’t there a law firm that hires people almost as good as the people on the mountain, for $100,000? This article suggests that “Not paying the standard top-tier salary [of $160,000] is a tacit admission that you’re no longer top-tier”, but you could say that about any industry where quality isn’t 100% obvious. How come chefs don’t have a salary graph that looks like that? How come engineers don’t?

It seems possible that maybe top law firms act as a de facto licensing system – picking out a couple of excellent young law school grads as Officially Excellent, and then if you’re a sufficiently big corporation you refuse to use any except those? But once again, I don’t know why law would develop this structure and other professions wouldn’t.

So if I had to figure out what all of these have in common, it would be an idea of privilege. Some people get guaranteed an unexpected privilege over and above the continuous measure of salary. The people who have to subsidize this privilege resent it and try to limit access to it. People start competing for scarce access to the privilege instead of having normal competitions for salaries, benefits, working conditions, et cetera, and all of those other things go out the window.

This is interesting because of how well it maps on to some other issues. For example, minimum wage creates a dualized system between workers and the unemployed. If there were no minimum wage, we would expect a sort-of-continuous wage distribution from 0.01$ an hour all the way up to whatever Taylor Swift makes for an hour’s performance. Instead, we guarantee everyone the privilege of $15 per hour. Employers resent this and (in theory) try to limit access to the privilege by lowering workforce, automating, etc, as much as possible. This creates a dualized system with an upper tier (employees with high wages) and a lower tier (unemployed with nothing at all).

Or how about benefits? If there were no benefits, we’d expect a more continuous spectrum of people working 40 hour weeks, 30 hour weeks, 20 hour weeks, and so on. Instead, we guarantee everyone who works X hours the privilege of good health care. Employers resent this and try to limit access to the privilege by hiring people to work X – 1 hours per week, or hiring independent contractors, or so on. This creates a dualized system with an upper tier (real employees) and a lower tier (people working 29.999 hours a week or whatever who don’t quite qualify for the benefits).

If you really want to stretch it, think about urban growth. If there was no zoning or regulation, desirable cities would have a continuous distribution from rich people living in nice mansions with lots of surrounding green land to poor people in apartment projects. Instead, we guarantee people living there certain privileges like “never having their view blocked” and “never having to worry about congestion”. This creates a two-tier system of current residents with the privileges, and non-residents who can’t live in desirable cities at all.

This raises a question of – assuming we want to give people privileges – or assuming we’re political realists who understand it’s going to happen anyway – are there ways to do it with a minimum of dualizing? It seems possible to imagine some solutions along those lines – for example, instead of mandating full health care for people who work more than 30 hours per week, we could seek systems where companies give health benefits that scale up with the number of hours worked. Instead of giving tenure, we seek systems where it becomes progressively harder to fire academics the longer they’ve worked for you.

Other cases seem harder – you can’t give half of a medical license to a doctor who finishes two years of med school, and the idea of a half a minimum wage defeats the whole point.

# Stalin and Summary Statistics

[Epistemic status: As always, I am not a statistician, and anything I say should be taken with a grain of salt until confirmed by others]

I.

A while ago, I wrote Beware Summary Statistics, where I talked about all the ways I’ve been misled by things like r-values and so on. I recently found some really interesting cases that brought up a few more some of these issues.

Back in June, Noah Smith blogged about a study on IQ And The Wealth Of States.

Some background: a group including Richard Lynn suggests that IQ is the driving factor behind income differences among countries. They are able to cite statistics on how a very rich country like Singapore has an average IQ of almost 110, and a very poor country like Haiti – well, it’s hard to say, because not too many Haitians take IQ tests and the ones who do might be so confused by this weird new idea of filling out a written multiple choice test that they choke and underperform – but officially Haiti has an IQ of like 70. Since you need smart people to build cool things like highways and power plants, maybe this explains a lot of the development/underdevelopment dichotomy. These people can point to a pretty good correlation between national IQ and national development to support their thesis, but the obvious counterargument is that maybe highly developed nations have good health and good education which raises IQ.

Anyway, the study Noah blogged about tested the application of this theory to US states. Noah sums up the results as follows:

The upper bound for the amount of state income differences that can be explained by population I.Q. differences is about a third. If we assume that achievement scores are a good measure of I.Q. and that school attainment doesn’t improve I.Q. very much, then the number goes down to about one-sixth.

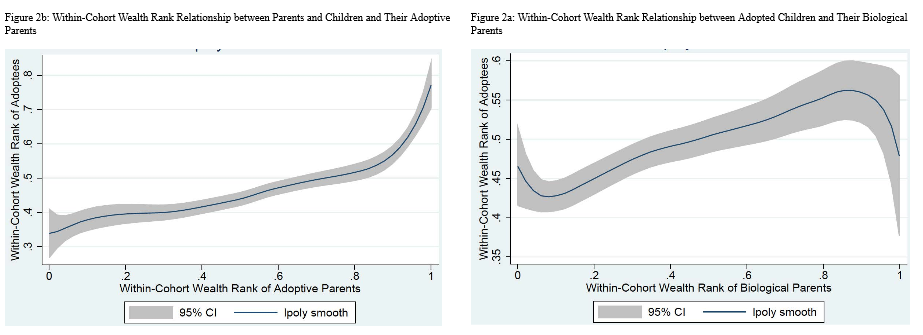
What this really shows is that there is Something Else that is driving state income differences. My personal guess is that this Something Else is mainly “external multipliers” from trade (the Krugman/Fujita theory). Institutions probably play a substantial role as well (the Acemoglu/Robinson theory). That’s certainly relevant for the debate about different models of capitalism, where we often compare the U.S. to Scandinavia and other rich places.

In any case, this result should be sobering for proponents of I.Q. as the Grand Unified Theory of economic development. Average I.Q. is not unimportant for rich countries, and we should definitely try to raise it through better nutrition, education, and (eventually) brain-boosting technologies. And it still might matter a lot for some poor countries. But for rich countries, there are things that matter a lot more.

II.

Let’s go to another study. The Atlantic has an article on How Rich People Raise Rich Kids, which is about Black, Devereux, Lundborg, and Majlesi (2015). They look at adopted Swedish kids and determine whether their wealth (not income!) as adults is more correlated with that of their biological parents or their adoptive parents. They find that non-adopted kids’ wealth correlates with that of their non-adopted parents at 0.33, adopted kids with biological parents at 0.13, and adopted kids with adoptive parents at 0.23. This suggests that upbringing is more important than genetics in determining how much wealth you will have.

Part of me wonders if an adoption study is really the best way to deal with this. Giving your children up for adoption is a very unusual choice, which means the biological parents are a very nonrepresentative group – and the study indeed finds that even forty years later, these biological parents have only a third as much money as the average Swede. If the same factors that cause them to give their children up for adoption – illness, relationship problems, trouble with the law – also cause them to fail to live up to their “genetic potential”, then we wouldn’t expect their children (who may lack these issues) to be correlated with them. The extremely odd shape of the graph also gives me pause: after a certain point, the wealthier your biological parents were, the less likely you are to be wealthy. Why? Certainly there’s no such effect for adoptive parents or non-adopted people!



But nitpicks aside, I am pretty willing to believe this. Although other studies have found evidence that biology is more important than upbringing in determining income (not wealth!), wealth seems like a different story. For one thing, you can just give your kids money! As I said last time we talked about GiveDirectly, there is pretty good evidence that giving people money causes them to in fact have the money which you just gave them. The current study reasonably tries to avoid having to deal with inheritances by looking at people whose parents are still alive, but even living parents can give lots of money to their children (for example, I come from a pretty wealthy family and my parents gave me lots of money, which I mostly used to help get through medical school without much debt. This means right now I have more “wealth” than people who took out bigger loans).

The authors write that:

While we have established the relative role of nature versus nurture, the exact mechanisms of wealth transmission are more deifficult to ascertain. Wealthier parents tend to be better educated and earn higher incomes, and these factors could lead to the increased wealth of their children through, for example, teaching them about investment opportunities or providing the right opportunities. However, when we investigate this, we find little evidence that this is the case. It may also be that wealthy parents invest more in their child’s education and career, which could then lead to higher child wealth accumulation. When we examine whether this is the case, however, we find little evidence for education or income as mechanisms. So the pathway through which parental wealth affects child wealth does not appear to be primarily parental schooling and income or child human capital accumulation and greater labor earnings. Taken together, our findings suggest potential roles for intergenerational transmission of preferences (children of wealthier parents may choose to save more or invest in assets that have higher returns) or for financial gifts from parents to children. Unfortunately, we do not have information on savings behavior or on financial gifts so this evidence is only suggestive.

So it seems to be a matter of how much money your parents give you, rather than of you learning deeply important personality traits from them or something. Fair enough.

But I got distracted. I was talking about the Atlantic’s article about the study. What did they have to say?

Even when they’re adopted, the children of the wealthy grow up to be just as well-off as their parents.

Lately, it seems that every new study about social mobility further corrodes the story Americans tell themselves about meritocracy; each one provides more evidence that comfortable lives are reserved for the winners of what sociologists call the birth lottery…What appears to matter—a lot—is environment, and that’s something that can be controlled.

III.

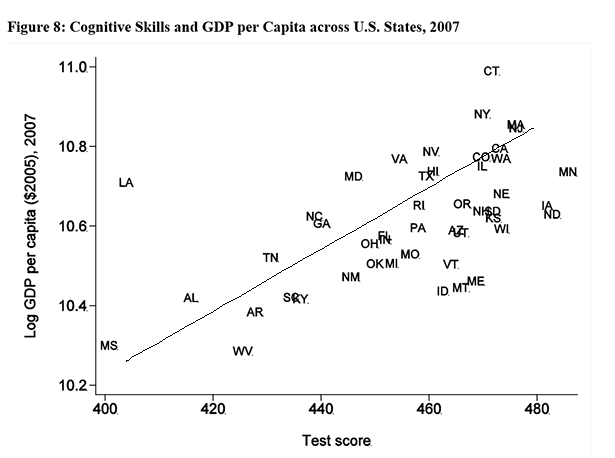
Let’s talk about three things – correlation, percent variance explained, and reality.

(I’m talking a big talk here, but I only got a good feeling for this when I asked various people on Tumblr to explain it to me. But they did a good job, and now I’m explaining it to you.)

Correlation is an r value. Percent variance explained is correlation squared. Reality is best viewed in the form of a graph.

Noah tells us that the IQ-of-states study found that only about 14% of the variation in state GDP was explained by IQ. Since variance = correlation^2, this implies that there’s a correlation of sqrt(0.14) = 0.37 between state IQ and state GDP. The paper itself did some sort of super high-powered nuclear statistics to arrive at this estimate, but I took lists of state average aptitude test scores and state GDP per capita and correlated them together in SPSS and got 0.40, so easy way and hard way agree pretty closely.

Here’s the graph associated with the study (I added the line):



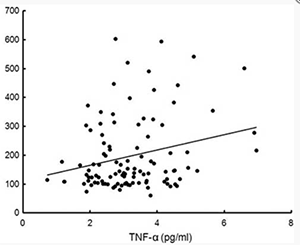
(Proposed new state motto: “Louisiana – Where We Succeed Wildly Out Of Proportion To Our Low Intelligence!”)

Huh! When you hear “…only explained 14% of the variance” it sounds like “go home, this is boring,” but when you hear “correlation of 0.37”, it sounds like “huh, they seem pretty related”, and when you see the graph, it looks like “holy frick, everything is IQ after all”. But all of these are the same finding!

Now. Consider the Swedish study and the Atlantic article about it. They say that although biological parents were correlated at r = 0.13, adoptive parents were correlated at r = 0.23. Therefore, they conclude, nurture wins over biology, meritocracy is a myth, everything depends on the lottery of birth, and wealthy parents are foredoomed to have wealthy children.

But r = 0.23 means the percent of variance explained is 0.23^2 = ~5%. If some Social Darwinist organization were to announce that they had evidence that who your parents were only determined 5% of the variance in wealth, it would sound like such overblown strong evidence for pure meritocracy that everyone would assume they were making it up.

The study didn’t come with a scatter plot, but here’s a plot from a totally different study that got a very similar correlation (0.24) to give you a feel for what it might look like:



The article makes it sound like your position in the birth lottery determines your destiny with impressive finality. The correlation seems unimpressive. The variance seems really unimpressive. The scatter plot looks like someone took random noise and drew a line through it. Once again, all the same finding.

Which of these three ways of presenting the data is most accurate? Um. Hard to say. I asked some people whether correlation (ie r = 0.23) or variance (ie 5%) is a better description of how the world actually works. That is, given that I have a certain “feel” for how much people differ in wealth, and a certain “feel” for what it means to win the birth lottery by getting rich parents, should I feel like the birth lottery thing explains 23% of how much wealth you have, or only 5%?

(I was only a Discordian for like six months, in my freshman year of college, but I still end up with fives and twenty-threes every time I try to do something involving numbers)

The answers I got were that it’s complicated, and both sort of work even though intuitively they should be mutually contradictory. The distribution of wealth is consistent with a story where it is explained by twenty different factors, each of which is just as important as parental wealth, which is sort of like 5%. But parental wealth explains just over a fifth of the standard deviation in wealth, which is sort of like 23%. The best explanation I got, from an anonymous commenter, was this:

About variance: consider the following. Flip 25 coins. Each heads gives you +1 utility point, and each tails gives you -1 utility point. One of these coins is labeled “upbringing”. On average you get 0 utility points. But you can also expect not to get exactly 0: on average, your distance from 0 will be 5 (the stdev is 5). So this is a little similar to a single coinflip that gives you either +5 or -5. Changing your upbringing from -1 to 1 gives you 2 points, out of a typical range of -5 to 5.

There was also a general consensus that if I had to think about this intuitively, which I should try not to do, 5% was probably the number that would lead me less astray, at least in terms of inputs. So fine. Whatever. Five percent it is.

IV.

Stalin once said that “The people who cast the votes decide nothing. The people who count the votes decide everything.”

(I briefly questioned whether Stalin really said that – like, I know he was an evil despot, but I’m not sure he was sufficiently self-aware about being an evil despot to come up with witty evil-despotism-related quotes. But I checked his WikiQuote page, and not only is the saying well-attested, but it seems Stalin was totally all about coming up with the witty self-aware evil-despotism-related quotes. Huh.)

In the same way, the people who conduct a study decide nothing. The people who report on the study decide everything.

I think Noah and the Atlantic were both honest and did a decent job reporting on their individual studies. But taken together, Noah concluded “This shows that IQ doesn’t really matter that much in explaining GDP” and the Atlantic concluded “This shows that who your parents are matters a colossally huge amount in explaining wealth” when in fact if you put both the studies side-by-side the IQ finding is three times as strong as the parents finding.

[EDIT: Some people have been misunderstanding this, so let me say it clearly. These are two studies about two different things! It’s like if I said the percent of weight gain explained by carbohydrates is three times as large as the percent of crime explained by poverty. I can compare these two things statistically, but I’m not trying to combine them into a single meta-study where I say that carbohydrates cause more crime than poverty! Also, some people seem to think I’m saying the Swedish study finds genes/biology/IQ to be more important than nurture. It doesn’t – in fact, it finds the opposite! Nurture is more important than genes but in the grand scheme of things both are tiny and the variance is almost entirely due to other things or randomness.]

In the end, nobody except a handful of researchers is going to remember the exact number. But they might remember “There were a couple interesting studies recently, one of them proved state IQ didn’t matter, the other proved that who your parents are totally determines whether you get ahead in life.” Framed that way, you might actually have gained negative knowledge from your diligent attempt to understand the economic literature.

And if the surrounding culture is pretty united in wanting to push a specific line, by choosing whether to publish r values or percent variance explained or graphs, they can pretty much hijack the intuitions even of people who don’t accept their reporting and try to rely on the numbers themselves.

The antidote is to have a good grasp of what each statistic means. And another antidote is to dial down your expectations. Remember, the study above was only able to correlate state IQ and state GDP at r = 0.4, but almost nothing in social science ever gets above 0.4. Trying to correlate rich parents with kids who become rich only got 0.2! 0.4 is pretty impressive and if you’re holding out for too much more you’re going to be living in a constant state of disappointment. I can think of one exception off the top of my head, and I am proud to say you will only find it here.

# Contra Hallquist On Scientific Rationality

I.

Topher Hallquist recently wrote a long article calling Less Wrong and the rationalist community “against scientific rationality” and having “crackpot tendencies”.

The piece claims to be about “the Less Wrong community”, but mostly takes the form of a series of criticisms against Eliezer Yudkowsky for holding beliefs that Hallquist thinks are false or overconfident. In some respects this is fair; Eliezer was certainly the founder of the community and his writings are extremely influential. In other respects, it isn’t; Margaret Sanger was an avowed eugenicist, but this is a poor criticism of Planned Parenthood today, let alone the entire reproductive rights community; Isaac Newton believed that the key to understanding the secrets of the universe lay in the dimensions of Solomon’s Temple, but this is a poor critique of universal gravitation, let alone all of physics. I worry that Hallquist’s New Atheism background may be screwing him up here: to critique a movement, merely find the holy book and prophet, prove that they’re fallible, and then the entire system comes tumbling to the ground. Needless to say, this is not how things work outside the realm of divine revelation.

On the other hand, it seems like the same argument that suggests Hallquist shouldn’t say such things would suggest I shouldn’t care much about arguing against them. I wish I lived in a universe where this was true, but “guilt by association” is a thing, the Internet has more than its share of people who have conceived this deep abiding hatred toward all rationalists, and “crackpot” and “anti-intellectual” are especially sticky accusations around these parts. Past experience tells me if I let this slide then at some point I’m going to be mentioning I’m interested in rationality and the automatic response will be “Oh, those are those anti-intellectual crackpots who hate science” and nothing I say will convince them that they are wrong, because why listen to an anti-science crackpot? Some things need to be nipped in the bud.

Also, a lot of Hallquist’s criticism is genuinely wrong and unfair. Also also, I like Eliezer Yudkowsky.

This is not to say that Eliezer – or anyone on Less Wrong – or anyone in the world – is never wrong or never overconfident. I happen to find Eliezer overconfident as heck a lot of the time. I have told him that, and he has pointed me to his essay on how if you really understand what confidence means in a probabilistic way, then you keep track of your uncertainty internally but don’t worry too much about the social niceties of broadcasting how uncertain you are to everyone. My opinion of this is the same as my opinion of most other appeals to not needing to worry about social niceties.

If Hallquist had made this reasonable critique, I would have endorsed it. Instead, I find his critique consistently misrepresents Eliezer, most of the ideas involved, and the entire Less Wrong community. I am going to fisk it hard, which I don’t like to do, but which seems like the only alternative to allowing these misrepresentations to stand. If you want to skip the (very boring) fisking, avoid part II and go straight from here to part III.

II.

So, to start with, I count four broad Eliezer-critiques:

1. Eliezer believes there’s an open-and-shut case for the Many Worlds Interpretation of quantum mechanics.  
2. Eliezer believes that some philosophical problems are really easy, and philosophers are morons for not having settled them already  
3. Eliezer is certain that cryonics will work  
4. Eliezer believes that “dietary science has killed millions of people” and cites borderline-crackpot Gary Taubes

A.

The first critique, about quantum mechanics, is potentially the strongest. Yudkowsky’s support for his preferred Many Worlds interpretation is remarkably strong for a subject where some of the world’s top geniuses disagree. On one hand, it is no stronger than that of many experts in the field – for example, Oxford quantum pioneer David Deutsch describes the theory’s lukewarm reception as “the first time in history that physicists have refused to believe what their reigning theory says about the world…like Galileo refusing to believe that Earth orbits the sun” and calls arguments against it “complex rationalizations for avoiding the most straightforward implications of quantum theory”. On the other, perhaps one could argue that a level of confidence appropriate in an Oxford professor is inappropriate in a self-taught amateur. I don’t know anything about quantum mechanics and don’t want to get into it.

Neither does Hallquist; he admits that Many Worlds is a reasonable position, but makes a different accusation. He says Yudkowsky has failed to inform readers to investigate other views:

Years ago in college I wrote a book debunking claims by Christian apologists claiming to “prove” that Jesus rose from the dead using historical evidence. At the very end of the book I included a little paragraph about how you shouldn’t just take my word for anything and should do your own research and form your own conclusions. In retrospect, that paragraph feels cheesy and obvious, but seeing the alternative makes me glad I included it.

Yudkowsky could have, after arguing at length for the many worlds interpretation of quantum mechanics, said, “I recommend going and studying the arguments of physicists who defend other interpretations, and when you do that I think you’ll see that physicists are screwing up.” That might have been reasonable. Many physicists accept many worlds, and I can accept that it’s sometimes reasonable for a dedicated amateur to have strong opinions on issues that divide the experts.

But in the very first post of his quantum physics sequence, Eliezer warns:

Everyone should be aware that, even though I’m not going to discuss the issue at first, there is a sizable community of scientists who dispute the realist perspective on QM. Myself, I don’t think it’s worth figuring both ways; I’m a pure realist, for reasons that will become apparent. But if you read my introduction, you are getting my view. It is not only my view. It is probably the majority view among theoretical physicists, if that counts for anything (though I will argue the matter separately from opinion polls). Still, it is not the only view that exists in the modern physics community. I do not feel obliged to present the other views right away, but I feel obliged to warn my readers that there are other views, which I will not be presenting during the initial stages of the introduction.

Okay, but that’s not exactly the same thing as telling his readers that they need to read other books and do other research and see whether he’s right or wr…

Go back and look at other explanations of QM and see if they make sense now. Check a textbook. Alternatively, check Feynman’s QED. Find a physicist you trust, ask them if I got it wrong, if I did post a comment. Bear in mind that a lot of physicists do believe MWI.

That’s from this comment. So Yudkowsky clearly did the thing Hallquist is accusing him of not doing.  
This is to his credit, but in my opinion supererogatory. Despite his anecdote about the paragraph in his book, Hallquist does not end his own essay on Less Wrong with “There are many people who disagree with me on this, be sure to check some pro-Less Wrong views to get the opposite side of the story.” Nor does he include such a sentence in the vast majority of his voluminous blog posts and writings. I don’t blame him for that – I don’t use such disclaimers either. There is a general conversational norm that to assert something is to say that you believe it and will provide evidence for your belief, not to say “You must believe me on this and may never go find any other sources to get any other side of the story.” You listen to me, you get my perspective, if for some reason I think you might not be aware that other perspectives exist I’ll tell you that they do, but I’m not going to end every blog post with “But be sure to read other sources on this so you can have your own opinion”.

Nevertheless, that is the standard Hallquist demands of Yudkowsky. And Yudkowsky meets it. And Hallquist ignores him doing so.

B.

Here’s philosophy. Note that I’m editing liberally to reduce length; I don’t think I’ve removed any essential parts of the argument but you might want to go over there and check:

In another post, Yudkowsky lodges a similar complaint about philosophy:

Philosophy is just not oriented to the outlook of someone who needs to resolve the issue, implement the corresponding solution, and then find out – possibly fatally – whether they got it right or wrong. Philosophy doesn’t resolve things, it compiles positions and arguments. And if the debate about zombies is still considered open, then I’m sorry, but as Jeffreyssai says: Too slow! It would be one matter if I could just look up the standard answer and find that, lo and behold, it is correct. But philosophy, which hasn’t come to conclusions and moved on from cognitive reductions that I regard as relatively simple, doesn’t seem very likely to build complex correct structures of conclusions.

I agree that progress in academic fields is sometimes slowed by the irrationality of the participants. People don’t like admitting being wrong. Unfortunately, knowing this isn’t much help unless you’ve discovered a magic formula for overcoming this flaw in human nature. More on that later. But thinking irrationality is the only reason why progress is slow ignores the fact that often, progress is slow because the questions are just really hard

[…]

The reason I’m saying all this is because when philosophers act like they’re not really trying to resolve debates, it’s because they know such attempts have a track record of not working. That doesn’t mean we will never put philosophy on a solid footing, but it does mean that anyone who shows up claiming to have done so single-handedly deserves a fair dose of skepticism.

The zombie debate is as good an example as any of this, so let’s talk about that. David Chalmers’ claim is that there could exist (in other possible worlds with different psychophysical laws–not the actual world) beings that are physically identical to us, but who lack consciousness. The intuition that such “zombies” are possible leads Chalmers to a view that at least looks a lot like epiphenomenalism (the belief in a separate mental realm affected by, but which does not affect, the physical realm).

Epiphenomenalism strikes a lot of people as crazy–me included! But Chalmers realizes this. So in The Conscious Mind he tries to do two things (1) argue his view is not quite epiphenomenalism (2) argue that some of the apparent advantages of certain other views over epiphenomenalism are illusory.

Does he succeed? I don’t know. But what makes me sympathetic to Chalmers is the sense that what he calls the hard problem of consciousness is a real problem, and alternative solutions aren’t any better. And Yudkowsky, as far as I can tell, isn’t one of those people who says, “the so-called ‘hard problem’ is a fake problem.” He agrees that it’s real–and then claims to have a secret solution he’ll sell you for several thousand dollars.

I think it’s enormously unlikely that Yudkowsky has really found the secret solution to consciousness. But even if he had, I don’t think anyone could know, including him. It’s like an otherwise competent scientist refusing to submit their work for peer review. Even top experts are fallible–and the solution is to have other experts check their work.

My response is going to sound mean and kind of ad hominem. That’s not really where I’m going here, and I promise I’m going somewhere a little more subtle than that. So, keeping that in mind – Hallquist once said on his blog:

I seem to have some philosophy-related skills in abundance. I’m good at spotting bad philosophical arguments. I was exposed to Plantinga’s modal ontological argument at ~12 years old, and instantly noticed it could just as well “prove” the existence of an all-powerful, necessarily existent being who wants nothing more than for everything to be purple.My experience with philosophy professors suggests that sadly, the knack for seeing through bad arguments is far from universal, even among the “professionals.”

What is the difference between Hallquist believing that he disproved one of the world’s most famous philosophers when he was twelve years old, and Eliezer believing that he solved the problem of consciousness when he was thirty-something?

Likewise, in another blog post Hallquist wrote about how Aquinas’ Arguments Just Suck and have no redeeming value:

Now, the broad point here is that while these arguments are supposedly derived from Aristotle, there doesn’t seem to be some secret Aristotelian assumption that would make them work. They’re just plain old bad arguments. I feel comfortable saying this, because respected living philosophers often give arguments that just stink, and being a contemporary of those philosophers I’m confident that the issue isn’t some peculiarly 21st century assumption.

Remember that Aquinas’ arguments convinced nearly all the brightest people in the Western world for five hundred years, and that many PhD philosophy professors still believe them with all their heart. Hallquist says they just suck and that there is no chance he might just be missing something or be misunderstanding their terms. In fact, he was so sure that there was no chance he was just a modern misunderstanding things that when a philosophy professor wrote a book claiming that modern people’s contempt for Aquinas is based on a misunderstanding, Hallquist read twenty pages of it, decided there was no chance that the remainder could possibly contain anything to change his mind, and stopped.

(I read the entire book, and if I’d stopped at page twenty I would have saved myself several valuable hours of my life I could have spent doing something productive, BUT THAT’S NOT THE POINT!)

Presumably, if Aquinas’ arguments are really stupid, but everyone believed them for five hundred years, this would imply there is something wrong with everyone. It might be worth saying “When someone makes a stupid argument, is there anything we can do to dismiss it in less than five hundred years?” But take that step, and you’re venturing into exactly the territory Hallquist is criticizing Eliezer for crossing into.

So what is my point beyond just an ad hominem attack on Hallquist?

Philosophy is hard. It’s not just hard. It’s hard in a way such that it seems easy and obvious to each individual person involved. This is not just an Eliezer Yudkowsky problem, or a Topher Hallquist problem – I myself may have previously said something along the lines of that anybody who isn’t a consequentialist needs to have their head examined. It’s hard in the same way politics is hard, where it seems like astounding hubris to call yourself a liberal when some of the brightest minds in history have been conservative, and insane overconfidence to call yourself a conservative when some of civilization’s greatest geniuses were liberals. Nevertheless, this is something everybody does. I do it. Eliezer Yudkowsky does it. Even Topher Hallquist does it. All we can offer in our own defense is to say, with Quine, “to believe something is to believe that it is true”. If we are wise people, we don’t try to use force to push our beliefs on others. If we are very wise, we don’t even insult and dehumanize those whom we disagree with. But we are allowed to believe that our beliefs are true. When Hallquist condemns Yudkowsky for doing it, it risks crossing the line into an isolated demand for rigor.

The most charitable I can be to Hallquist is that his gripe is not with Yudkowsky’s hubris in holding strong beliefs, but with an apparent reluctance to publish them. Once again, from the end of the quoted material:

What makes me sympathetic to Chalmers is the sense that what he calls the hard problem of consciousness is a real problem, and alternative solutions aren’t any better. And Yudkowsky, as far as I can tell, isn’t one of those people who says, “the so-called ‘hard problem’ is a fake problem.” He agrees that it’s real–and then claims to have a secret solution he’ll sell you for several thousand dollars.

I think it’s enormously unlikely that Yudkowsky has really found the secret solution to consciousness. But even if he had, I don’t think anyone could know, including him. It’s like an otherwise competent scientist refusing to submit their work for peer review. Even top experts are fallible–and the solution is to have other experts check their work.

So perhaps the difference between Hallquist (and the rest of us) and Yudkowsky is that the latter doesn’t believe in peer review and openness because he “hasn’t published” his “secret” solution to consciousness?

Hallquist’s only source for Eliezer having such a solution is the Author Notes for Harry Potter and the Methods of Rationality, Chapter 98, where somewhere in the middle he says:

I am auctioning off A Day Of My Time, to do with as the buyer pleases – this could include delivering a talk at your company, advising on your fiction novel in progress, applying advanced rationality skillz to a problem which is tying your brain in knots, or confiding the secret answer to the hard problem of conscious experience (it’s not as exciting as it sounds).

Even if, as Hallquist claims, he has private information that this isn’t a joke, this hardly seems like the central case of Eliezer having a philosophical opinion. One might ask: has Eliezer ever written about his other philosophical beliefs, the ones that seem most important to him?

The Less Wrong Sequences are somewhere between 500,000 and a million words long (for comparison, all three Lord of the Rings books combined are 454,000, and War and Peace is 587,000). Eliezer may be one of the most diligent people alive about publicizing his philosophical opinions in great detail. In some cases, he explicitly frames his Sequence work in terms of getting feedback – for example, in the same comment on the Quantum Physics Sequence I linked earlier, he writes about his approach to double-checking his quantum work:

I myself am mostly relying on the fact that neither Scott Aaronson nor Robin Hanson nor any of several thousand readers have said anything like “Wrong physics” or “Well, that’s sort of right, but wrong in the details…”

And this works – far from being unwililng to debate academic philosophers about his opinion of Chalmers’ epiphenomenalism, he argues the subject with David Chalmers himself in the comments of that Less Wrong post, which is a heck of a lot more than I’ve ever done when I disagree with an academic philosopher.

On the other hand, since he placed a single sentence in an HPMOR Author’s Note about a solution to the hard problem of consciousness which he hasn’t written about, Hallquist accuses him of being against publicizing his work for review or criticism. Once again, I think these complaints are startlingly unfair and a remarkably flimsy ground on which to set out to tarnish the reputation of an entire community.

C.

Hallquist’s next complaint is that Eliezer is strongly in favor of cryonics; he starts by noting that Eliezer criticizs Michael Shermer for mocking the role of molecular nanotechnology in cryonics, then writes:

Full disclosure: I’m signed up for cryonics. But the idea that nanomachines will one day be able to repair frozen brains strikes me as highly unlikely. I think there’s a better chance that it will be possible to use frozen brains as the basis for whole brain emulation, but I’m not even sure about that. Too much depends on guesses both about the effects of current freezing techniques and about future technology.

Eliezer, meanwhile, is sure cryonics will work, based, as far as I can tell, on loose analogies with computer hard drives. Faced with such confident predictions, pointing out the lack of evidence and large element of wish-fulfillment (as Shermer does) is an eminently reasonable [reaction]. A “rationalism” that condemns such caution isn’t worthy of the name.

I can’t stress enough the enormous difference between trying to do some informed speculation about what technologies might be possible in the future, and thinking you can know what technologies will be possible in the future based on just knowing a little physics. Take, for example, Richard Feynman’s talk “There’s Plenty of Room at the Bottom”, often cited as one of the foundational sources in the field of nanotechnology.

Today, the part of Feynman’s talk about computers looks prophetic, especially considering the talk was given several years before Gordon Moore made his famous observation about computer power doubling every couple of years. But other things he speculates about are, to say the least, a long way off. Do we blame Feynman for this?

No, because Feynman knew enough to include appropriate caveats. When he talks about the possibility of tiny medical robots, for example, he says it’s “very interesting possibility… although it is a very wild idea.” He doesn’t say that this will definitely happen and be the secret to immortality. And some futurists, like Ray Kurzweil, do say things like that. That’s the difference between having a grasp of the difficulty of the topic and your own fallibility, and well, not.

Yet Yudkowsky is so confident in his beliefs about things like cryonics that he’s willing to use them as a reason to distrust mainstream experts.

I hate to keep getting into these tu quoque style arguments. But I do feel like once you have signed up for cryonics, you lose the right to criticize other people for being crackpots for being signed up for cryonics for a slightly different reason than you are. If you think molecular nanotechnology is highly unlikely but whole brain emulations aren’t, please at least be aware that from the perspective of everyone else in society you are the equivalent of a schizophrenic guy who believes he’s Napoleon making fun of the other schizophrenic guy who thinks he’s Jesus.

But Hallquist’s main criticism isn’t just that Yudkowsky believes in cryonics, or nanotechnology, or whatever. It’s that he’s overconfident in these things. As per the quote, he’s “sure cryonics will work” and says the equivalent of “this will definitely happen and be the secret to immortality”.

At this point, it will probably come as no surprise that Yudkowsky has never said anything of the sort and has in fact clearly said the opposite.

One of my favorite results from the Less Wrong Survey, which I’ve written about again and again, shows that people who sign up for cryonics are less likely to believe it will work than demographically similar people who don’t sign up (yes, you read that right) – and the average person signed up for cryonics only estimated a 12% chance it would work. The active ingredient in cryonics support is not unusual certainty it will work, but unusual methods for dealing with moral and epistemological questions – an attitude of “This only has like a 10% chance of working, but a 10% chance of immortality for a couple of dollars a month is an amazing deal and you would be an idiot to turn it down” instead of “this sounds weird, screw it”. Once you think of it that way, signing up doesn’t mean “I’m sure it will work” but rather “I’m not 100% sure it won’t.” Thus, Eliezer’s former co-blogger Robin Hanson, who frequently joined forces with Eliezer in passionate appeals for their mutual readers to sign up for cryonics, gives his probability of cryonics working at 6%.

I don’t have as accurate a picture of Eliezer’s beliefs. The best I can do is try to interpret this comment, where he describes the success of cryonics as a chain of at least three probabilities. First, the probability that the core technology works. Second, the probability that cryonics organizations don’t go bankrupt. Third, the probability that humankind lasts long enough to develop appropriate resurrection technology.

Eliezer gives the first probability as 80 – 90%, the second probability as “outside the range of my comparative advantage in predictions”, and the third probability as “the weakest link in the chain” and as “something I’ve refused to put a number on, with the excuse that I don’t know how to estimate the probability of doing the ‘impossible'”. So his probability of cryonics working seems to be 80-90% \* something-he-doesn’t-know \* something-he-thinks-is-near-impossible. If we are willing to totally ignore his request not to put numbers on these, perhaps something like 90% \* 75% \* 10%? Which would land him right between Robin Hanson’s 6% and the survey average of 12%.

I think Eliezer’s overconfident on his belief on the first of his three bins, the one about the core technologies working. But Hallquist ignores this relatively modest issue and instead chooses to sensationalize Eliezer into being “sure cryonics will work” and “[thinking] this will definitely happen”, when his actual probability is unknown but probably closer to 6% than 100%. Once again I think this complaint is startlingly unfair and a remarkably flimsy ground on which to set out to tarnish the reputation of an entire community.

[my own disclaimer: I am not signed up for cryonics and don’t currently intend to, but I respect people who are]

D.

Oh God, we’re going to have to get into diet again. You can skip this part. Seriously, you can.

Eliezer writes that “dietary scientists ignoring their own experimental evidence have killed millions and condemned hundreds of millions more to obesity with high-fructose corn syrup” and links to this article by Taubes.

I do not want to defend Gary Taubes. Science has progressed to the point where we have been able to evaluate most of his claims, and they were a mixture of 50% wrong, 25% right but well-known enough that he gets no credit for them, and 25% right ideas that were actually poorly known enough at the time that I do give him credit. This is not a bad record for a contrarian, but I subtract points because he misrepresented a lot of stuff and wasn’t very good at what might be called scientific ethics. I personally learned a lot from reading him – I was able to quickly debunk the wrong claims, and the correct claims taught me things I wouldn’t have learned any other way. Yudkowsky’s reading of him seems unsophisticated and contains a mix of right and wrong claims. But Hallquist’s reading seems to be a prime example of reversed stupidity not being intelligence. He writes:

The use of the diet example is even more embarrassing than the other claims I’ve looked at so far. The line about “dietary scientists ignoring their own experimental evidence” links to an article by Gary Taubes. Taubes champions the diet claims of Robert Atkins, who literally claimed that you could eat unlimited amounts of fat and not gain weight, because you would pee out the excess calories. This, needless to say, is not true.

After reading two of Taubes’ books, I haven’t been able to find anywhere where he addresses the urine claim, but he’s very clear about claiming that no amount of dietary fat can cause weight gain. How Taubes thinks this is supposed to be true, I have no idea. His attempted explanations are, as far as I can tell, simply incoherent. (Atkins at least had the virtue of making a coherent wrong claim.)

I’m startled by Hallquist’s claim that he read two Taubes books and couldn’t find Taubes’ explanation for why people don’t gain calories on a high-fat diet. Taubes lays out this mechanism very clearly as the major thesis of his book.

Taubes believes the human body is good at regulating its own weight via the hunger mechanism. For example, most Asian people are normal weight, despite the Asian staple food being rice, which is high-calorie and available in abundance. Asians don’t get fat because they eat a healthy amount of rice, then stop. This doesn’t seem to require amazing willpower on their part; it just happens naturally.

In a similar vein is one of Taubes’ favorite studies, the Vermont Prison Experiment, where healthy thin prisoners were asked to gain lots of weight to see if they could do it. The prisoners had lots of trouble doing so – they had to force themselves to eat even after they were full, and many failed, disgusted by the task. Some were able to eat enough food, only to find that they were filled with an almost irresistible urge to exercise, pace back and forth, tap their legs, or otherwise burn off the extra calories. Those prisoners who were able to successfully gain weight lost it almost instantly after the experiment was over and they were no longer being absolutely forced to maintain it. The conclusion was that healthy people just can’t gain weight even if they want to, a far cry from the standard paradigm of “it takes lots of willpower not to gain weight”.

Other such experiments focused on healthy thin rats. The rats were being fed as much rat food as they wanted, but never overate. The researchers tried to trick the rats by increasing the caloric density of the rat food without changing the taste, but the rats just ate less of it to get the same amount of calories as before. Then the researchers took the extreme step of surgically implanting food in the rats’ stomachs; the rats compensated by eating precisely that amount less of normal rat food and maintaining their weight. The conclusion was that rats, like Asians and prisoners, have an uncanny ability to maintain normal weight even in the presence of unlimited amounts of food they could theoretically be binging on.

Modern Westerners seem to be pretty unusual in the degree to which they lack this uncanny ability, suggesting something has disrupted it. If we can un-disrupt it, “just eat whatever and let your body take care of things” becomes a passable diet plan.

I sometimes explain this to people with the following metaphor: severe weight gain is a common side effect of psychiatric drug Clozaril. The average Clozaril user gains fifteen pounds, and on high doses fifty or a hundred pounds is not unheard of. Clozaril is otherwise very effective, so there have been a lot of efforts to cut down on this weight gain with clever diet programs. The journal articles about these all find that they fail, or “succeed” in the special social science way where if you dig deep enough you can invent a new endpoint that appears to have gotten 1% better if you squint. This Clozaril-related weight gain isn’t magic – it still happens because people eat more calories – but it’s not something you can just wish away either.

Imagine that some weird conspiracy is secretly dumping whole bottles of Clozaril into orange soda. Since most Americans drink orange soda, we find that overnight most Americans gain fifty pounds and become very obese.

Goofus says: “Well, it looks like Americans will just have to diet harder. We know diets rarely work, but I’m sure if you have enough willpower you can make it happen. Count every calorie obsessively. Also, exercise.”

Gallant says: “The whole problem is orange soda. If you stop drinking that, you can eat whatever else you want.”

Taubes’ argument is that refined carbohydrates are playing the role of Clozaril-in-orange-soda. If you don’t eat refined carbohydrates, your satiety mechanism will eventually go back to normal just like in Asians and prisoners and rats, and you can eat whatever else you want and won’t be tempted to have too much of it – or if you do have too much of it, you’ll exercise or metabolize it away. When he says you can “eat as much fat as you want”, he expects that not to be very much, once your broken satiety mechanism is fixed.

Taubes is wrong. The best and most recent studies suggest that avoiding refined carbohydrates doesn’t fix weight gain much more than avoiding any other high-calorie food. However, the Clozaril-in-orange-soda model, which is not original to Taubes but which he helped popularize, has further gained ground and is now arguably the predominant model among dietary researchers. It’s unclear what exactly the orange soda is – the worst-case scenario is that it’s something like calorically-dense heavily-flavored food, in which case learning this won’t be very helpful beyond current diet plans. The best-case scenario is that it’s just a disruption to the microbiome, and we can restore obese people to normal weight with a basic procedure which is very simple and not super-gross at all.

But whether or not you agree with it, this Clozaril-in-orange-soda story is indisputably Taubes’ model; Hallquist seems to miss this and instead makes vague gestures towards a discredited 70s theory that calories are excreted as ketones in the urine.

Because Hallquist doesn’t understand Taubes’ main point, his criticisms miss the mark. He wrote a five-part Less Wrong series on Gary Taubes, in which he tries to figure out what Taubes’ theory of obesity is and as best I can tell somehow ends up simultaneously saying Taubes is dishonest for accusing mainstream researchers of thinking diet is a matter of willpower, and saying Taubes is silly because diet really is just a matter of willpower. If you don’t believe me, read the post, where he says:

Taubes goes on at great length about how obesity has other causes beyond simple calorie math as if this were somehow a refutation of mainstream nutrition science. So I’m going to provide a series of quotes from relevant sources to show that the experts are perfectly aware of that fact.

Which he does. But then later he says:

So what’s going on here? I think the answer lies Taubes’ eagerness to portray mainstream nutrition experts as big meanies who blame fat people for being fat…

But this puts Taubes in a bind: now if he says how much we eat has an effect on our weight, he’s a big meanie too. It doesn’t work for him to say fat people can’t help overeating because of something wrong with their metabolism, and this in turn causes them to gain weight, because he’s committed himself to the principle that blaming behavior equals blaming a character defect. So instead, we get wild rhetoric about how stupid the experts are with no coherent view underneath it.

A more sensible approach would’ve been to emphasize that akrasia is an extremely common problem for humans, and that people who don’t suffer from akrasia in regards to diet probably suffer from akrasia about something else. But that wouldn’t have made for as an exciting of a book.

So unless I am reading this wrong, he thinks the correct answer is to say that we should blame behavior, but that we should do it in a nice way where we say kind things about how everybody has willpower problems and it’s nothing to be ashamed of. I think he thinks he’s agreeing with mainstream consensus here, but mainstream consensus has already moved on to “Screw willpower, STOP DRINKING ORANGE SODA”.

Next, Hallquist attacks Taubes’ claim that “We don’t get fat because we overeat; we overeat because we’re getting fat”, saying that he’s “trying to be charitable” but “surely it wasn’t meant to be taken literally” and that he’s “playing with meanings” just so he can “portray nutrition experts as big meanies”.

But when you ask those nutrition experts – for example, Dr. David Ludwig MD PhD Professor of Nutrition at Harvard, they’re writing articles in the Journal of the American Medical Association with titles like Increasing Adiposity – Cause Or Consequence Of Overeating?, where they say things like:

Since [the early part of the century], billions of dollars have been spent on research into the biological factors affecting body weight, but the near-universal remedy remains virtually the same, to eat less and move more. According to an alternative view, chronic overeating represents a manifestation rather than the primary cause of increasing adiposity. Attempts to lower body weight without addressing the biological drivers of weight gain, including the quality of the diet, will inevitably fail for most individuals…a focus on total diet composition, not total calories, may best facilitate weight loss.

And although the journal article is relatively balanced and might be dismissed as just a scholarly investigation of a weird ideas, the same authors wrote up the same argument for the New York Times in a more obviously persuasive fashion.

Hallquist, again thinking he’s defending a consensus position, attacks Gary Taubes’ claim that government guidelines promoting low-fat diets were responsible for the increase in sugar and refined carbohydrate consumption. He calls this “a huge red flag”, said Taubes is “unaware of [reality] or tries to hide it from his readers”, engages in “irresponsible rhetoric” and wonders “how anyone reading this could avoid suspecting something was up.”

Meanwhile, let’s go to more experts writing in the Journal of the American Medical Association – we’ll keep Ludwig but add Dr. Dariush Mozaffarian, MD, MPH, Dean of Tufts University School of Nutrition – explaining why the new dietary guidelines have done a sudden about-face, removed previous restrictions on fat, and added more restrictions on refined carbohydrates:

With these quiet statements, the DGAC report reversed nearly 4 decades of nutrition policy that placed priority on reducing total fat consumption throughout the population. In 1980, the Dietary Guidelines recommended limiting dietary fat to less than 30% of calories. This recommendation was revised in 2005, to include a range from 20% to 35% of calories. The primary rationale for limiting total fat was to lower saturated fat and dietary cholesterol, which were thought to increase cardiovascular risk by raising low-density lipoprotein cholesterol blood concentrations. But the campaign against saturated fat quickly generalized to include all dietary fat. Because fat contains about twice the calories per gram as carbohydrate or protein, it was also reasoned that low-fat diets would help prevent obesity, a growing public health concern.

The complex lipid and lipoprotein effects of saturated fat are now recognized, including evidence for beneficial effects on high-density lipoprotein cholesterol and triglycerides and minimal effects on apolipoprotein B when compared with carbohydrate. These complexities explain why substitution of saturated fat with carbohydrate does not lower cardiovascular risk.

Most importantly, the policy focus on fat reduction did not account for the harms of highly processed carbohydrate (eg, refined grains, potato products, and added sugar)—consumption of which is inversely related to that of dietary fat…Based on years of inaccurate messages about total fat, a 2014 Gallup poll shows that a majority of US residents are still actively trying to avoid dietary fat, while eating far too many refined carbohydrates.

Hallquist reserves special mockery for Taubes’ claim that the government’s low-fat mania reached such intensity that nutritional initiatives promoted soda:

This portrayal of mainstream nutrition science is as false as Atkins’ claim about peeing out excess calories. Besides the obvious – who on earth ever believed Coca-Cola was a health food?

Once again, the United States’ top nutrition scientists, explaining the new consensus dietary guidelines in the New York Times:

The “We Can!” program, run by the National Institutes of Health, recommends that kids “eat almost anytime” fat-free salad dressing, ketchup, diet soda and trimmed beef, but only “eat sometimes or less often” all vegetables with added fat, nuts, peanut butter, tuna canned in oil and olive oil. Astoundingly, the National School Lunch Program bans whole milk, but allows sugar-sweetened skim milk. Consumers didn’t notice, either. Based on years of low-fat messaging, most Americans still actively avoid dietary fat, while eating far too much refined carbohydrates.

Hallquist critiques Yudkowsky’s worry dietary scientists have been too soft on high-fructose corn syrup. Once again, the latest dietary guidelines almost halve the allowed amount of HFCS from the permitted amount in the last set of guidelines.

Hallquist derides Yudkowsky’s claims that dietary science has “killed millions”. An article by one of Britain’s top doctors, in the British Medical Journal, which by the way is the third highest-impact medical journal in the world, asks Are Some Diets Mass Murder? and argues that

[A] consequence of the fat hypothesis is that around the world diets have come to include much more carbohydrate, including sugar and high fructose corn syrup, which is cheap, extremely sweet, and “a calorie source but not a nutrient.” More and more scientists believe that it is the surfeit of refined carbohydrates that is driving the global pandemic of obesity, diabetes, and non-communicable diseases. They dispute the idea that we get fat simply because energy in exceeds energy out, saying instead that the carbohydrates “trigger a hormonal response that drives the portioning of the fuel consumed as storage as fat… The successful attempt to reduce fat in the diet of Americans and others around the world has been a global, uncontrolled experiment, which like all experiments may well have led to bad outcomes

If indeed there were serious flaws in the dietary guidelines for the past thirty years, and since obesity kills about 370,000 people per year, if the issues corrected in the latest guidelines and freely admitted by modern scientists made the problem even 10% worse, then the “millions of deaths” figure is not an exaggeration.

Taubes is wrong about a lot of things. There is a lot of room for someone to criticize Taubes, and indeed, I have done so repeatedly. Hallquist tries to criticize Taubes, but fires so indiscriminately that he manages to reserve some of his strongest condemnation for claims of Taubes which are actually true and widely recognized as such.

Hallquist writes of Yudkowsky:

Remember, this is one of Yudkowsky’s go-to examples for why you shouldn’t trust the mainstream too much! And it’s not just wrong, it’s wrong in a way that could have been caught through common sense and basic fact-checking. But I guess common sense is just tribalistic bias, and who needs fact-checking when you’ve got superior rationality? The nicest thing you can say about this is that, when he encourages his followers to form strong opinions based on the writings of a single amateur, he’s only preaching what he practices.

I am usually in favor of being nice to people who get things wrong, because things are hard and goodness knows I am wrong often enough. But I am not in favor of being nice to people who get things wrong and are smug and mean to everyone else about them, because punishing defectors is the only way things ever get done in this world. So:

Topher. You seriously do not understand Taubes. You somehow read his book while by your own admission missing the entire mechanism he was trying to explain. You then go on to call a bunch of propositions ludicrous, idiotic, not-even-wrong, et cetera, when those propositions are widely acknowledged as true by the scientific community you think you are defending. You are nevertheless setting yourself up as an expert and trying to explain these subjects to other people. Many of these people told you this when you first posted on LW, and you ignored those people and keep trying to do it. Mozaffarian, Ludwig, Friedman, etc are the United States’ top nutritional scientists, and they are telling you this. I am telling you this. Everyone is telling you this, and you are putting your fingers in your ears and shouting “EVERYTHING IS SO OBVIOUS, I CAN’T BELIEVE OTHER PEOPLE GOT THIS WRONG, IT’S ALL SO EASY, EVERYONE EXCEPT ME IS AN IDIOT.”

Eliezer Yudkowsky has had some pretty silly ideas about diet. I know this because he when he has them, he comes to me and asks me if they are correct, and I tell him. At one point, he bought and sent me a book he was interested in so that I could review it and tell him if it made sense. I told him it was wrong, and he listened. If you had asked me if you were right in your criticisms of Taubes, I would also have reviewed them and explained them to you. You didn’t, because you were so certain that you had to be right that you didn’t need to consult with anybody else, despite the fact that you are an amateur with no medical knowledge.

Thus is it written: “Why do you look at the mote in Eliezer Yudkowsky’s eye, and ignore the beam in your own?”

The last I heard about Eliezer’s dietary philosophy was his OKCupid profile, where under “Food” he wrote: “Flitting from diet to diet, searching empirically for something that works.”

SUCH OVERCONFIDENCE. SO CERTAINTY. VERY ANTI-SCIENCE.

III.

Okay, now that I’ve gotten my nitpicks out of the way, what about the actual meat of Hallquist’s criticism?

Hallquist claims that Less Wrong is fundamentally anti-science. All of his criticisms of Eliezer Yudkowsky were to show examples of him behaving in anti-science ways, but he also thinks that Eliezer comes right out and admits it:

Now that I’m thousands of words and about as many tangents into this post, let me circle back to something to something I said early in the post: pointing out the flaws in mainstream experts only gets you so far, unless you actually have a way to do better. This isn’t an original point. Robin Hanson has made it many times. (See here for just one example.) But I want to emphasize it anyway.

It’s the main reason I’m unimpressed with the material on LessWrong about how the rules of science aren’t the rules an ideal reasoner would follow. This is a huge chunk of Yudkowsky’s “Sequences”, but suppose that’s true, so what? We humans are observably non-ideal. Throwing out the rules of science because a hypothetical ideal reasoner wouldn’t need them is like advocating anarchism on the grounds that if Superman existed, we’d have no need for police.

I think this is more than a superficial analogy. To borrow another point from Hanson, most of us rely on peaceful societies rather than personal martial prowess for our safety. Similarly, we rely on the modern economy rather than personal survival skills for food and shelter. Given that, the fact that science is, to a large extent, a system of social rules and institutions doesn’t look like a flaw in science. It may be the only way for mere mortals to make progress on really hard questions.

Yudkowsky is aware of this argument, and his response appears to mostly depend on assuming the reader agrees with him that physicists are being stupid about quantum mechanics–that, combined with a large dose of flattery. “So, are you going to believe in faster-than-light quantum ‘collapse’ fairies after all? Or do you think you’re smarter than that?” asks one post.

This is combined with an even stranger argument, an apparent belief that it should be possible for amateurs to make progress faster than mainstream experts simply by deciding to make progress faster. Remember how the imagined future “master rationalist” complains “Eld scientists thought it was acceptable to take thirty years to solve a problem”? This is a strange thing to complain about. Either you have a way to make progress quickly or you don’t, and if you don’t, you don’t have much choice but to accept that fact.

Back in the real world, wishing away the difficulty of hard problems don’t make them stop being hard. This doesn’t mean progress is impossible, or that’s it’s not worth trying to improve on the current consensus of experts. It just means progress requires a lot of work, which most of the time includes first becoming an expert yourself, so you have a foundation to build on and a sense of what mistakes have already been made. There’s no way to skip out on the hard work by giving yourself superpowers.

I agree that you don’t make progress faster just by “wishing away the difficulty” or “giving yourself superpowers” or “deciding to make progress faster”.

On the other hand, if Yudkowsky thought that becoming more rational was a matter of “wishing away the difficulty”, he wouldn’t have written a larger-than-Lord-of-the-Rings introduction to the subject. He would have just wished.

Developing and learning an Art Of Thinking Clearly isn’t just “wishing away the difficulty” of settling on true ideas faster, any more than developing and learning rocket science is “wishing away the difficulty” of going to the moon. Thinking clearly is super-hard, but perhaps it is a learnable skill.

Rocket science is a learnable skill, but if you want to have it you should probably spend at least ten years in college, grad school, NASA internships, et cetera. You should probably read hundreds of imposing books called things like Introduction To Rocket Science. It’s not something you just pick up by coincidence while you’re doing something else.

If thinking clearly is a learnable skill, where are the grad schools for it? Where are the textbooks? Not in philosophy programs – Hallquist and I both agree about that. What all of this “only domain-specific knowledge stuff matters” effectively implies is that “thinking clearly” is so easy you can pick it up by coincidence while working on pretty much anything else – something we believe about practically no other skill. If you trusted a rocket scientist who had never read a single rocket science textbook to be any good at rocket science, you’d be insane, but we routinely trust the subjects we most need to think clearly about to people who have never read a How To Think Clearly textbook – and I can’t blame us, because such textbooks, or at least good evidence-based textbooks of the same quality as the rocket science ones, simply don’t exist.

The Sequences aren’t an assertion that you can wish away a problem. They’re a cry for textbooks.

But Hallquist has a counterargument:

The big difference between what [scientists] do and what Yudkowsky advocates is that probability theory is much less useful here than a good knowledge of cell biology.

If we want to get all hypothetical, we can imagine some kind of theorizing contest between a totally irrational person with an encyclopaedic knowledge of cell biology, versus a very rational person who knows nothing at all about the subject. Who would win? Well, who cares? Whoever wins, we lose. We lose because I want the people working on curing cancer to be good at both cell biology and thinking clearly, to know both the parts of science specific to their own field and the parts of science that are as broad as Thought itself. I have seen what happens when people know everything about cell biology and nothing about rationality. You get AMRI Nutrigenomics, where a bunch of people with PhDs and MDs give a breathtakingly beautiful analysis of the complexities of the methylation cycle, then use it to prove that vaccines cause autism. By all means, know as much about methylation as they do! But you’ve also got to have something they’re missing!

I want people who know as much about the methylation cycle as the Nutrigenomics folks, while also understanding the idea of privileging the hypothesis. I don’t want to defy experts, I want to give experts better tools.

In fact, even that framing isn’t quite right. Every day I have patients come to me and ask questions like “are benzodiazepines safe and effective?” or “is therapy better than SSRIs?” or “will this drug increase my risk of dementia?” or “does untreated bipolar increase my risk of converting to rapid-cycling?” or a host of other questions. And I ask my mentor, who’s one of the top psychiatrists in the state, and he gives me a nice, straightforward answer, and then I ask my mentor at the other hospital I go to, and he’s also one of the top psychiatrists in the state, and he gives me precisely the opposite answer. And when I mention to either of them that the other guy disagrees, they just assure me that if I do the research myself I’ll find that their point of view is obviously and self-evidently correct. And meanwhile, my patients are pressing me for answers and telling me that if I get this wrong it will ruin their life. And I can’t say “Wait fifty years until enough studies are done to be totally sure.”

“Don’t worry too much about learning rationality, just listen to the experts” is all nice and well up until the point where someone hands you a lab coat and says “Congratulations, you’re an expert!” And then you say: “Well, frick.” And when that day comes you had better already have learned something about the Art Of Thinking Clearly or else you have a heck of a steep learning curve ahead of you.

Hallquist says that Less Wrong is “against scientific rationality”. Well, we’re “against scientific rationality” in the same sense that my hypothetical Soviet who says “We need two Stalins! No, fifty Stalins!” is against Stalinism as currently implemented. It is in the right direction. But it needs to go further. This is why all of the posts Hallquist finds to support his assertion that Less Wrong is “against scientific rationality” are called things like Science Isn’t Strict Enough. I’m “against scientific rationality” insofar as when my patients demand answers to semi-impossible questions and say their lives depend on it, I want to have scientific rationality on my side, and another tool, and a third tool if I can think of it, and as many extra tools as it takes before I stop being terrified.

If you don’t trust the quantum mechanics sequence to make the point for you – and maybe you shouldn’t – I explain my own version of this revelation in the highly-Eliezer-inspired The Control Group Is Out Of Control. Science is what hands us an unusually well-conducted meta-analysis proving that psi exists with p < 1.2 \* 10^-10, crowning fifty years of parapsychological research that finds positive results about as often as not. Bayes is what tells us that parapsychology makes no sense, has an ungodly level of Kolmogorov complexity, and is going to require a heck of a lot more than a good meta-analysis before we accept it. In that sense, "switching allegiance from Science to Bayes" isn't some cataclysmic event where we foresake Galileo thrice before an onyx altar, it's something we all do already under the right circumstances. The point is figuring out how to formalize it, so that we don't mess up and dismiss a result that's counterintuitive but true. I respect Yudkowsky's decision not to use an example like this because if he used this example people would assume he was only talking about parapsychology and real science is totally safe, but I think he was going for the same principle.

I have immense respect for Topher Hallquist. His blog has enlightened me about various philosophy-of-religion issues and he is my go-to person if I ever need to hear an excruciatingly complete roundup of the evidence about whether there was a historical Jesus or not. His commitment to and contribution to effective altruism is immense, his veganism puts him one moral tier above me (who eats meat and then feels bad about it and donates to animal charities as an offset), and his passion about sex worker rights, open borders, and other worthy political causes is remarkable. As long as Topher isn’t talking about diet or Eliezer Yudkowsky’s personal qualities, I have a lot of trust in his judgment.

But these things I like and respect about Topher are cases where he’s willing to go his own way. He views open borders as an pressing moral imperative even though you’ll have a hard time finding more than a handful of voters, sociologists, or economists who support it. He’s signed up for cryonics even though 99% of the population think that makes him crazy. He donated to fight AI risk way back when it was hard to find any AI experts willing to endorse the cause, and so gains extra credibility and moral authority now that many of them have. Heck, I even respect his ability to put down a terrible Aquinas book on the twenty-somethingth page when I trudged all the way through.

And – and this is a compliment, so I hope he takes it as one – I wish he would try to help spread his own good qualities. We need more people who are able to evaluate difficult moral and intellectual arguments and come to apparently-bizarre but in-fact-very-important conclusions, even when there is not a knock-down scientific argument proving them correct quite yet.

And a necessary consequence of having people who are able to go beyond the things that have knock-down scientific proofs, and go beyond the things that everyone by consensus agrees to be true, and who are able to discuss weird ideas like effective altruism and cryonics and the Singularity, is that occasionally some people will venture too far and say something genuinely out of line (remember: decreasing your susceptibility to Type I errors will always increase your susceptibility to Type II errors, and vice versa!) When this has happened in the rationality community, I have tried again and again to politely but firmly correct these people.

I would like to have Topher as a partner in this effort, but instead, I find him to be trawling the entire corpus of everything people in the rationalist community have ever said or done for quotes he can take out of context to “prove” that they are “crackpots” and that they universally “hate experts”. It’s led to him rushing through books he doesn’t really understand so he can get to the fun part where he points out how crackpotty everyone else is for not rejecting the book fast enough. It’s led to him gradually burning bridges with a lot of people who should be on his side by being needlessly hostile to them. It’s led to him turning Yudkowsky’s opinion that science needs to be stricter and stronger into Eliezer being “against scientific rationality” and “anti-intellectual” and “pro-crackpot”, peppered with a laundry list of out-of-context gripes. It’s not a productive way to have the discussion and, more importantly, it’s not true. And it’s not fair to the efforts that the rationalist community keeps putting in to improve themselves and their thought processes.

If Eliezer Yudkowsky ever showed up and said “I have perfected this Art, now I am never wrong,” then I would happily join Hallquist in laughing hysterically.

If Eliezer Yudkowsky showed up and said “I have tiny pieces of this Art and some promising leads on who can help us find more, let’s work on it together,” well, I’ve spent the past couple of years taking that offer and so far I don’t regret it.

And if Eliezer Yudkowsky showed up and said “I thought I had pieces of the Art, but I was wrong, I don’t know anything about it, nobody does,” then I will still go to my grave believing that whether or not we know it, such an art should exist, that even if it’s near-impossible we should be chipping away at the impossibility as much as we can in the hopes of getting a couple of tiny shards of something useful that we can cherish as precious.

But I think it isn’t as bad as all that. We do have some tiny preliminary seeds of such an Art. I think such an art involves learning to appreciate your cognitive biases on a gut level. I think it involves understanding the relevant basics of probability theory and calibration. I think it involves knowing when to use the Inside View or the Outside View, how to avoid getting bogged down in meaningless semantic arguments, and how to overcome your resistance to changing your mind in the face of new evidence. It also involves knowing how to read studies, learning to get a feel for the process of science and find out who is and isn’t a credible expert, learning when science does and doesn’t work and how to repair the latter category, learning to avoid the well-known pitfalls, and learning how to build communities where good epistemology can flourish.

It also involves a bunch of other things that I don’t know and Eliezer doesn’t know and maybe no one in our community knows, but once we find out, we intend to steal them, and you should help.

# Contrarians, Crackpots, and Consensus

I.

Last week we discussed whether Gary Taubes gets to be admitted to the small but prestigious pantheon of correct contrarians. And the strange part was that there was a lot less argument about how correct he was than about how contrarian he was.

Taubes’ main theory – that low-carb diets could solve the obesity epidemic – hasn’t fared the test of time very well. But some of his supporting points have. Large parts of mainstream nutrition science have eased up on dietary cholesterol, dropped the recommendation against fat, gotten tougher on sugar, and accepted that the science should focus on how to regulate complex satiety mechanisms rather than just counting calories. Given how hard it is to fight the scientific consensus and win, even those few minor victories would potentially be remarkable.

The counterargument is that these are other people’s ideas and he gets no credit for them. Suppose David Icke says that the Queen is a lizard person, and also that the royal family is secretly descended from German nobles and isn’t British at all. A hundred years from now his readers celebrate his genius: although he got the lizard part wrong, the Saxe-Coburg-Gotha theory was remarkably prescient!

If Icke’s book spends just as much time arguing for the Saxe-Coburg-Gotha theory as for the lizard theory, all while inveighing against some supposed consensus of anti-Saxe-Coburg experts, we can imagine the far future finding him pretty impressive, since they might not have done the historical work necessary to realize everyone knew of the Queen’s foreign origins all along. Even though the Queen’s German descent sounds shocking, and even though the average British person probably isn’t aware of it, and even though it’s something nobody really likes to talk about – doesn’t mean Icke “discovered” it in any interesting way. He’s not prescient, he just sometimes reads Wikipedia in between his bizarre ravings.

Thing is, even though the 1990s were like twenty years ago and pretty well-documented, people have had a surprisingly hard time coming to agreement on how novel Taubes’ ideas were back then.

They certainly weren’t perfectly novel – Taubes himself tries to claim he’s just relaying ideas from scientists and researchers to the public, and even his most controversial theories come from other people like Dr. Atkins. And they certainly weren’t perfectly well-known – everyone has a story of their doctor or dietician or something telling them “just eat low-fat foods, cut back on cholesterol, and count calories”. But there seems to be a lot of room in between those two poles.

This is starting to remind me of another debate I got stuck in recently – my argument with Rob Wipond about the serotonergic theory of depression. Wipond argued that psychiatrists irresponsibly promoted a narrative in which depression was a simple serotonin deficiency and so taking Prozac would quickly and elegantly solve the problem. I told him that actually, no, the psychiatric community wasn’t saying that at all, which was why every single example he thought he could find of that turned out to be a garbled out-of-context quote which when investigated honestly was clearly saying the opposite. I got a lot of angry comments that no, people were very sure their doctor had told them that depression was a simple serotonin deficiency.

I think a lot of things are getting obscured by the term “scientific establishment” or “scientific consensus”. Imagine a pyramid with the following levels from top to bottom:

FIRST, specialist researchers in a field. So for example the people doing studies on the effect of dietary cholesterol, or the people dissecting monkey brains to see how much serotonin is in them. These people always have the latest cutting-edge experimental results and a good knowledge of the issues involved in the field.

SECOND, non-specialist researchers in a broader field. Nutrition scientists in general. The guy who is interested in Vitamin B, but goes to the same conferences as the guys studying cholesterol. The research psychiatrist working on schizophrenia, but who maintains a keen interest in what her colleagues over in the depression lab are doing. They know enough about the broad principles of the field to be able to understand and evaluate new ideas more quickly than everybody else, but they still only learn about them the same way everyone else does – by waiting for the specialist researchers to tell them.

THIRD, the organs and administrators of a field who help set guidelines. The head of the USDA who’s in charge of looking over the Food Pyramid to make sure it’s accurate. The APA Committee for deciding exactly what wording to use in the guidelines on depression treatment. The head of Harvard Medical School who has to decide what to put in the curriculum. The editor of the New England Journal of Medicine, who has to decide what gets published.

FOURTH, science journalism, meaning everyone from the science reporters at the New York Times to the guys writing books with titles like The Antidepressant Wars to random bloggers.

ALSO FOURTH IN A DIFFERENT COLUMN OF THE PYRAMID BECAUSE THIS IS A HYBRID GREEK PYRAMID THAT HAS COLUMNS, “fieldworkers”, aka the professionals we charge with putting the research into practice. In nutrition this is doctors and dieticians, who directly inform their patients what to eat. In education research this could be teachers and principals who directly decide how classes will get taught. In sociology it might be the police chief trying to institute a new crime-fighting program. Et cetera.

FIFTH, the general public.

A lot of these issues make a lot more sense in terms of different theories going on at the same time on different levels of the pyramid. I get the impression that in the 1990s, the specialist researchers, the non-specialist researchers, and the organs and administrators were all pretty responsible about saying that the serotonin theory was just a theory and only represented one facet of the multifaceted disease of depression. Science journalists and prescribing psychiatrists were less responsible about this, and so the general public may well have ended up with an inaccurate picture.

Likewise, when Taubes published his book, the ideas he wrote about (at least the correct ones) seem to have been accepted by some specialist researchers, known only as vague inklings among non-specialist researchers, poorly reflected at all in the official dietary guidelines, totally new to the world of journalism, totally new to doctors (who mostly still haven’t gotten the message), and totally new to the general public.

This whole process gets even more complicated when you consider enemy action. In psychiatry, drug companies have established defensive chokeholds at various points on the pyramid, trying to promote pro-pharmaceutical results and sink anti-pharmaceutical ones. This isn’t a far-out conspiracy theory – practically every psychiatrist agrees it’s true to some degree, which is why there are so many conflict-of-interest laws to try to minimize the damage. The only debate is whether we’ve successfully contained it to a small effect, or whether it’s hopelessly contaminated the entire process (I tend to lean more toward the optimistic side; for a true pessimist, read Dr. Nardo). The same is true in nutrition, where a lot of studies are sponsored by groups with names like ‘The United Dairy Farmers Council’ or ‘The League For Wheat’. Even when there aren’t official lobbyists, political opinion plays a big part: the social science journals are full of studies that very competently show that certain politically popular ideas are bunk; by the time they reach the ears of voters and policymakers this has mysteriously been transformed into “scientists agree with you that these politically popular ideas need much more funding”. When you have a block in the process like this, the specialist researchers, the non-specialists, the guideline-makers, the fieldworkers, and the public can all remain on totally different pages for a surprisingly long time.

Taubes – and some of the people making the most noise about the serotonin theory of depression – seem to be people trying to transfer knowledge from the highest levels of the pyramid all the way down to the base, skipping over the levels in between. Does that make them contrarians playing Galileo to a hidebound establishment, or responsible science journalists relaying the establishment’s ideas more faithfully and efficiently than their predecessors? Your opinion probably depends on what narrative suits your purpose at any given time.

II.

I think of some of the contrarians who seem to have their heads screwed on straight. Irving Kirsch and Robert Whitaker on antidepressants. Cochran and Harpending on recent human evolution. Judith Rich Harris on parenting. Nick Bostrom on superintelligence.

I don’t agree with all these people, I’ve even written long rants against some of them. But they seem to be of a different breed than crackpots like creationists and parapsychologists and anti-vaxxers. It’s hard to specify how. It’s not just credentialed expertise. Michael Behe and Daryl Bem are both professors, and Andrew Wakefield was an MD who’d done previous published immunological research, but their work falls squarely in the ‘crackpot’ column.

But one thing I do notice about these virtuous contrarians – their reception is surprisingly quiet. We know creationism is wrong partly because half the evolutionary biologists in the world have written books about why creationism is wrong, which they advertise prominently on their blogs about why creationism is wrong. Where are all the developmental psychologists shouting down Judith Rich Harris? I’ve seen a few very specialized psychiatrists argue against Kirsch, but never very heatedly, and usually while granting many of his points. The majority of the profession? Never heard of him and don’t care.

When I first became interested in AI risk around 2007, people told me that no legitimate AI experts were seriously worried. I checked and at the time that was mostly true. On the other hand, no legitimate AI experts were specifically not worried either. AI risk just wasn’t their area, and they were perfectly happy to ignore it and concentrate on things that were. There are two types of “no evidence”, and this was the entirely neutral one. It seemed like a very different situation than vaccines causing autism. There, too, experts in the field aren’t worried – but they’re not worried because every single one of them has an opinion and the opinion is “NO”.

(Sure enough, since then a lot more AI researchers have become interested, in exactly the sort of sea change I don’t expect to see mirrored in the autism field.)

The crackpots seem to be met with violent opposition. The virtuous contrarians seem to be met with – well, almost boredom. No one is particularly interested in adopting their ideas, but no one is particularly interested in arguing against them either.

(on the other hand, Time Cube Guy is met with boredom by serious astronomers, either because he’s too small to be noticed or too small to be worth refuting, so it’s not like it’s a great heuristic)

A while ago, I was reading some stuff about the role of choline in the brain, and I thought: “I wonder if anyone has ever used this to treat bipolar disorder”. Well, I searched the literature, and there was one very small study from 1996 in which choline apparently demonstrated excellent effect treating rapid-cycling bipolar disorder, which is otherwise quite difficult to treat. The study isn’t obscure – it seems to have been cited 110 times – but no one’s followed up on that and you could easily go your whole life studying psychiatry without running into any kind of choline-bipolar connection. It seems like a potentially important idea, which has small but nonzero evidence behind it, but which everyone nevertheless ignores, because it isn’t anybody’s business in particular. There are hundreds of things like this scattered across the literature in pretty much every field.

If I were to announce that small-minded scientists were ignoring the result of their own research and covering up The Truth About Choline, possibly at the behest of lizard-people…well, I could certainly do it in a crackpottish way if I wanted to. But I’d be interesting. I wouldn’t be trivially wrong in the same sense as the homeopath who doesn’t care that all biologists disagree with them.

Thomas Kuhn categorized scientific progress into everyday advances and “paradigm shifts”, the latter being major reconceptualizations like the one between geocentrism and heliocentrism, or from classical mechanics to quantum mechanics.

If I understand right – everyone is doing science, and occasionally they come up with something that doesn’t make sense. Whatever. Half the time people come up with things that don’t make sense, and it usually just means your neutrino speedometer is miscalibrated. They either refuse to publish, because no point in publishing nonsense, or they publish, everyone says “Huh, that’s funny”, and they continue doing what they’re doing. The view of science presented to students in the field, and the one that the luminaries of the field think in most of the time, is the one made up of all of the nice consistent results that make sense, with the noise abstracted away.

Then somebody looks a little closer and sees a pattern in the noise. The studies and ideas everyone else was ignoring actually tell a consistent story which is more plausible than the grand narrative of the field which everyone else is working off of. They propose a new paradigm. There is some fighting and eventually it is determined to be superior to the old one, which is jettisoned in its favor.

Someone does a study on Tibetans and says “it looks like they’re adapted to their mountain environment, but that would require really fast evolution, which we all know practically never happens, so it’s probably some weird fluke.” Someone does a study on Indo-Europeans and says “it looks like they have unique lactose tolerance, but that would require really fast evolution, which we all know practically never happens, so it’s probably some weird fluke.” Someone does a study on Ashkenazi Jews and says “it looks like they have higher-than-average intelligence, but that would require really fast evolution, which we all know practically never happens, so it’s probably some weird fluke.” Then Cochran and Harpending and a few others take a sweeping view of everything, and say “OR WHAT IF REALLY FAST EVOLUTION IS HAPPENING PRETTY MUCH ALL THE TIME?!” They’re not exactly pulling this discovery ex nihilo, but they’re taking what might be the private opinion of a couple of isolated specialist researchers who might not have known one another, synthesizing all the evidence together, and saying the thing nobody else wanted to mention.

One way to be a contrarian without being a crackpot seems to be trying to start these sorts of paradigm shifts. Indeed, I notice that they are often people with enough expertise to understand a field who nevertheless acquired that expertise outside of the field itself. For example, Kirsch is a psychologist, as opposed to the psychiatrists and biochemists who usually deal with antidepressant drugs. Cochran is a physicist by training, although he somehow ended up as an anthropology professor. Harris was pursuing a psychology PhD but quit for health reasons and did most of her research independently. Bostrom is a philosopher, and so has license to stick his finger in pretty much whatever pots he wants.

(Whitaker is a journalist. So is Gary Taubes – and, for that matter, Steve Sailer. Science journalism seems like a good example of how somebody can learn a lot about a field while still having an outsider perspective on it)

At their best, these people can look at a field, find ideas that have been excluded from the narrative, and create a new narrative around them. Sometimes this goes horribly wrong – this is how I think of Graham Hancock, also a journalist, who took every weird archaeological discovery and mysterious ancient monument and fit them together into a brilliant, wacky, but ultimately completely bonkers narrative of ancient supercivilizations. It also seems to be how Taubes blundered into his low-carb fanaticism.

So this can sort of be a red flag. But it’s a much less glaring red flag than when people like homeopaths or anti-vaxxers believe they have discovered a new effect, and continue to maintain it exists despite real scientists’ insistence that it doesn’t.

And these are the people who are most likely to get caught in the trap mentioned in Part I. If they’re doing their job right, all they’re doing is calling increased attention to certain results in the field. They’re not the first people to mention that there’s some evidence for recent human evolution. They might not even be the first people to publish a review paper collecting a bunch of different examples of recent human evolution in one place. They’re the first people to be jerks about it, the first people to say “HEY, YOU WITH THE PARADIGM, YOU SUCK” and force all the lower levels of the pyramid – the non-specialists, the administrators, the fieldworkers, the journalists, and the public – to confront the new possibilities head-on.

But shouting “YOU SUCK” doesn’t win anybody any friends. Even if their side triumphs in the end, there will be many much more sober academics who were pushing it almost as effectively. And the same perversity of spirit that led contrarians to challenge the field where it was wrong will probably make them overshoot and challenge the field where it is right. Thus, Taubes not only says that fat has been unfairly demonized, but goes further and says that fat is great for you and you can stay at whatever weight you want just by eating fat. Kirsch and Whitaker not only say that antidepressants were less effective than previously believed, they say they’re worthless for most people and will poison you and psychotherapy is great. Judith Rich Harris not only says that quirks of parenting style don’t matter, she also minimizes the effects of divorce – which I think goes too far.

The likely outcome is pretty much what we’ve got. Even when contrarians win, they lose. Members of the field will be celebrated for being the ones who helped usher in the new paradigm. And the contrarians will be remembered as partisans of crazy false ideas, who happened to gain a thin veneer of credibility by also repeating some true stuff already known by domain experts.

(I’m very happy that brilliant AI researchers like Stuart Russell have joined the fight against AI risk. But I fully expect future textbooks to say that Russell is a great hero for discovering AI risk single-handedly, and also there were some weird guys in Berkeley who gained superficial credibility by parroting Russell’s theories, but were really just silly people writing fanfiction.)

John Baez’s Crackpot Index offers thirty points for “fantasizing about show trials in which scientists who mocked your theories will be forced to recant.”

And if you think you’re a true genius who will have the last laugh, the joke’s on you. You won’t get your show trials even if you’re right.

# Book Review: Chronicles Of Wasted Time

I.

I was recently recommended Chronicles of Wasted Time, the autobiography of Malcolm Muggeridge. It was a good choice, and not just because its title appropriately described my expectations about reading 500-page books on people’s recommendation. Muggeridge is an obvious reactionary, but one with the personal and historical credentials to pull it off with the utmost class and credibility.

He describes his birth in 1903 to a family of committed British socialists. Their heroes were Karl Marx, George Bernard Shaw, and Fabian leaders Sidney and Beatrice Webb. These last two I had only the slightest familiarity with, but Muggeridge paints a picture of them as the progressive titans of his day, boasting a combination of Chomsky’s intellectual leadership and the Clintons’ network and political acumen. Throughout Muggeridge’s youth, his family would host meetings, sing socialist songs, run for various minor offices on the socialist ticket, and exchange correspondence with intellectual worthies. They even flirt with, though never quite join, an experimental commune being set up in their area, about which Muggeridge has the best stories:

The land was cheap in those days, and they acquired it by purchase; then, to demonstrate their abhorrence of the institution of property, ceremonially burnt the title deeds. It must have been a touching scene – the bonfire, the documents consigned to the flames, their exalted sentiments. Unfortunately, a neighboring farmer heard of their noble gesture and began to encroach on their land. To have resorted to the police, even if it had been practicable, was unthinkable. So after much deliberation, they decided to use physical force to expel the intruder; which they did on the basis of a theory of detached action, whereby it is permissible to infringe a principle for the purpose of a single isolated act without thereby invalidating it. The intruding farmer was, in fact, thrown over the hedge in the presence of the assembled Colonists. There were many such tragi-comic incidents in the years that followed; as well as quarrels, departures, jealousies, betrayals, and domestic upsets. In the end, the Colonists found it necessary to reestablish their title to the land by means of squatters’ rights, and then proceeded to bicker amongst themselves as to who should have which portion.

But he and his family are convinced that all of this is just a momentary hiccup on the road to Glorious Progress. Indeed, his teenage years are marked by a burning excitement at the Russian Revolution:

We called the Metropolitan Mounted Police ‘Cossacks’, rejoiced over early Soviet films like ‘Mother’ and ‘The Battleship Potemkin’, spoke of workers’ control and cadres and agitprop, and I personally decided inwardly that sooner or later I would go to Russia and throw in my lot with the new and better way of life that, I was confident, was coming to pass there.

Against this enthusiasm, he had only a personal tendency which he describes as a deep-set conviction:

…that I was born into a dying, if not already dead, civilization, whose literature was part of the general decomposition; a heap of rubble scavenged by scrawny Eng Lit vultures, and echoing with the hyena cries of Freudians looking for their Marx and Marxists looking for their Freud…a Gaderene descent down which we all must slide, finishing up in the same slough.

By the same token, a strange certainty has possessed me, almost since I can remember, that the Lord Mayor riding in his coach, the Lord Chancellor seated on his Woolsack, Honorable and Right Honorable Members facing one another across the floor of the House of Commons, were somehow the end of a line. That soon there would be no more Lord Mayors, Lord Chancellors, Honorable and Right Honorable Members, the Mother of Parliaments having reached her time of life or menopause, and become incapable of any further procreation…

Doubtless other glories lie ahead. Bigger and better capsules carried to the moon; down in the test tube something stirs; ‘I think, therefore you’re not’ says the computer. We all know, though, in our hearts, that our old homestead is falling down; with death-watch beetles in the rafters, and dry rot in the cellar, and unruly tenants whose only concern is to pull the place to pieces.

This feeling – that everything around him was in a state of permanent decay – was not so far-fetched given that he spent much of his early adulthood in the far-flung territories of the crumbling British Empire. But it soon becomes clear that it’s more than a natural reaction to the political realities of the time. He describes again and again looking on something apparently healthy enough and being overwhelmed with a feeling of impending sickness and decay. He describes T.S. Eliot as “a death-rattle in the throat of a dying civilization”, Shaw as “too encased in his own narcissism, too remote from real life to do more than grimace at it through a long-distance telescope”, and the great reformers and abolitionists of the age as:

…solemn funeral mutes in the long obsequies of western civilization; as they fell by the way, others coming forward to take their places. Now the time has nearly come for the coffin to be actually interred. Then at last their occupation will be gone forever.

I sometimes have patients with very severe depression who tell me that everything they look at is infested by maggots. They won’t eat, because the food is infested with maggots. They won’t hug their children, because their children are infested with maggots. Sleep disgusts them, because the bed is infested with maggots. Et cetera.

And other times, when they have a little more insight, they’ll say something like “Okay, my food isn’t literally infested by maggots, but I get this feeling from it, this overwhelming feeling, such that the feeling would only make sense if the food was infested by maggots. I know deep down it’s not infested by maggots, but it has some metaphysical quality which only things infested by maggots have.”

Poor Malcolm Muggeridge feels this way about everything. One of the most poignant episodes in the book takes place the worst night of the London Blitz, when Muggeridge runs around the burning city, almost euphoric, because finally his inner conviction that everything is on fire and collapsing is reflected in everything really being on fire and collapsing, and nobody can pat his head and patronizingly tell him that it isn’t:

I remember particularly Regent’s Park on a moonlit night, full of the fragrance of the rose gardens; the Nash Terraces, perfectly blacked-out, not a sign of a light anywhere, white stately shapes waiting to be toppled over – as they duly were, crumbling into rubble like melting snow…I felt a terrible joy and exaltation at the sight and smell and taste and sound of all of this destruction; at the lurid sky, the pall of smoke, the faces of bystanders wildly lit in the flames. Goebbels, in one of his broadcasts, accused us of glorying obscenely in London’s demolition. He had a point, but what he failed to understand was that we had destroyed our city already before the Luftwaffe delivered their bombs; what was burning was no more than the dry, residual shell.

The only things that seem to give him any kind of brief reprieve from the maggots are church services, classic literature, quiet domestic life with his wife and 2.4 children, and rural country fields.

And he is convinced, absolutely convinced, that he should be a socialist and go move to the USSR.

This goes approximately as well as you would expect.

After graduating college, which he dislikes because maggots, he gets a couple of jobs at various far-flung British Empire outposts, which he hates. Then, somewhat by coincidence, he ends up in journalism.

His reaction to journalism is an increasing terror that this might be his calling. He is very good at it, takes to it like an old veteran almost immediately, feels in some strange way that he has come home – but the entire enterprise fills him with loathing. He watches in horror how easily the words flow on to the page when his puppet-masters tell him to argue for a particular cause, how fluidly he takes to idioms like “It is surely incumbent upon all of us to…” and “there can be no one here present who does not…”. He writes:

So I began, and the words seemed to come of themselves; like lying as a child, or as a faithless lover; words pouring out of one in a circumstantially false explanation of some suspicious circumstance. The more glib, the greater the guilt…it is painful to me now to reflect the ease with which I got into the way of using this non-language; these drooling non-sentences conveying non-thoughts, propounding non-fears and offering non-hopes. Words are as beautiful as love, and as easily betrayed. I am more penitent for my false words – for the most part, mercifully lost forever in the Media’s great slag-heaps – than for false deeds.

But Malcolm Muggeridge isn’t going to take this lying down! Malcolm Muggeridge has a plan! Malcolm Muggeridge is going to escape this duplicitous charade of lies and petty propaganda. Malcolm Muggeridge is going to move to Stalin’s USSR.

So he does.

He gets a job as The Guardian‘s Russia correspondent and sets off for Moscow with a host of other British intellectuals, heading for what all of them expect is the Promised Land. The mood on their ship is electric; he describes them all singing, sure that they are leaving behind this wretched bourgeois world for the Golden Future:

On their way to the USSR they were in a festive mood; like a cup-tie party on their way to a match, equipped with rattles, coloured scarves and favors. Each of them harboring in his mind some special hope; of meeting Stalin, or alternatively, of falling in with a Komsomolka, sparkling eyed, red scarf and jet black hair, dancing the carmagnole, above all, with very enlightened views on sex, and free and easy ways…oh, to be in Russia, now that Stalin’s there!

His excitement dissipates relatively early; he finds that the Soviet journalistic world fails to live up to his expectations:

Being a correspondent in Moscow, I found, was, in itself, easy enough. The Soviet press was the only source of news; nothing happened or was said until it was reported in the newspapers. So all I had to do was go through the papers, pick out any item that might be interesting to readers of the Guardian, dish it up in a suitable form, get it passed by the censor at the Press Department, and hand it in at the telegraph office for dispatch. One might, if in a conscientious mood, embellish the item a little…sow in a little local colour, blow it up a little, or render it down a little according to the exigencies of the new situation. The original item itself was almost certainly untrue or grotesquely distorted. One’s own deviations, therefore, seemed to matter little, only amounting to further falsifying what was already false.

This bizarre fantasy was very costly and elaborate and earnestly promoted. Something gets published in Pravda; say, that the Soviet Union has a bumper wheat harvest – so many poods per hectare. There is no means of checking; the Press Department men don’t know, and anyone who does is far, far removed from the attentions of foreign journalists. Soviet statistics have always been almost entirely fanciful, though not the less seriously regarded fro that. When the Germans occupied Kiev in the 1939-45 war they got hold of a master Five Year Plan, showing what had really been produced and where. Needless to say, it was quite different from the published figures. This in no way affected credulity about such figures subsequently, as put out in Russia, or even in China.

Hey man, don’t knock China, they’re doing great! Their GDP rose 7% this year! It must be true! The Guardian tells us so!

But getting back to the story…although it is clear to him that the Soviet economy is struggling, every dispatch they are given to send home declares that things are better than ever, that the Workers’ Paradise is even more paradisiacal than previously believed, that the evidence is in and Stalinism is the winner. It doesn’t matter what he makes of this, because anything he writes which deviates from the script is rejected by the censors, who ban him from sending it home. He is reduced to sending secret messages at the bottoms of people’s suitcases, only to find to his horror that even when they successfully reach the Guardian offices back in Britain, his bosses have no interest in publishing them because they offend the prejudices of its progressive readership. Finally, he finds himself a part of the elite fraternity of western journalists on the Soviet beat, who maintain their morale by one-upping each other in how cynical and patronizing they can be towards their Russian hosts and their credulous readers back home:

We used to run a little contest among ourselves to see who could produce the most striking example of credulity among this fine flower of our western intelligentsia. Persuading church dignitaries to feel at home in an anti-God museum was too easy to count. So was taking lawyers into the people´s courts. I got an honourable mention by persuading Lord Marley that the queueing at food shops was permitted by the authorities because it provided a means of inducing the workers to take a rest when otherwise their zeal for completing the five-year plan in record time was such that they would keep at it all the time, but no marks for floating a story that Soviet citizens were being asked to send in human hair – any sort – for making of felt boots. It seemed that this had actually happened.

And he remembers the contempt of these grizzled veterans for the steady stream of Western tourists, intellectuals, and general Stalin fanboys who arrived to gawk over the Glorious New Civilization:

I have never forgotten these visitors, or ceased to marvel at them, at how they have gone on from strength to strength, continuing to lighten our darkness, and to guide, counsel and instruct us. They are unquestionably one of the wonders of the age, and I shall treasure till I die as a blessed memory the spectacle of them travelling with radiant optimism through a famished countryside, wandering in happy bands about squalid, over-crowded towns, listening with unshakeable faith to the fatuous patter of carefully trained and indoctrinated guides, repeating like schoolchildren a multiplication table, the bogus statistics and mindless slogans endlessly intoned on them. There, I would think, an earnest office-holder in some local branch of the League of Nations Union, there a godly Quaker who had once had tea with Gandhi, there an inveigher against the Means Test and the Blasphemy Laws, there a staunch upholder of free speech and human rights, there an indomitable preventer of cruelty to animals, there scarred and worthy veterans of a hundred battles for truth, freedom, and justice – all, all chanting the praises of Stalin and his Dictatorship of the Proletariat. It was as though a vegetarian society had come outwith a passionate plea for cannibalism, or Hitler had been nominated posthumously for the Nobel Peace Prize.

His final break with the rest of the enlightened progressive world comes when he decides to do something that perhaps no other journalist in the entire Soviet Union had dared – to go off the reservation, so to speak, leave Moscow undercover, and see if he can actually get into the regions where rumors say some kind of famine might be happening. The plan goes without a hitch, he passes himself off as a generic middle-class Soviet, and he ends up in Ukraine right in the middle of Stalin’s Great Famine. He describes the scene – famished skeletons begging for crumbs, secret police herding entire towns into railway cars never to be seen again. At great risk to himself, he smuggles notes about the genocide out of the country, only to be met – once again – with total lack of interest. Guardian readers don’t look at the newspapers to hear bad things about the Soviet Union! Guardian readers want to hear about how the Glorious Future is already on its way! He is quickly sidelined in favor of the true stars of Soviet journalism, people like Walter Duranty, the New York Times‘s Russia correspondent, who wrote story after story about how prosperous and happy and well-fed the Soviets were under Stalin, and who later won the Pulitzer Prize for his troubles.

Muggeridge, on the other hand, penurious from lack of interest in his stories, fearing for his safety from the Soviet government, and generally disgusted with everything – even more so than usual for a world infested with maggots – decides to get the hell out of Dodge. He’s had enough of Russia, enough of Communism, enough of that entire part of the world. He’s going somewhere safe, somewhere decent. He’s going somewhere that will renew his crumbling faith in humanity. He’s going to Nazi Germany right as the anti-Jewish pogroms are starting.

Well, to make a long story short, this doesn’t restore his faith in humanity. He hangs out in Berlin for a while, sending his pieces on the Russian famine to all the newspapers he knows, watching more and more rejections come in each day, earning the ire of all of his leftist friends for apparently deserting the cause and turning traitor. Finally, he tells his boss:

“From the way you’ve cut my messages about the Metro-Vickers affair, I realize that you don’t want to know what’s going on in Russia, or let your readers know. If it had been an oppressed minority, or subject people valiantly struggling to be free, that would have been another matter. Then any amount of outspokenness, any amount of honesty.”

I went on to describe the scene in Berlin, and the Nazis beating up Jewish shops, and everyone with his story of murder and folly, and concluded:

“It’s silly to say the Brown Terror is worse than the Red Terror. They’re both horrible. They’re both Terrors. I watched the Nazis march along Unter den Linded and realized – of course, they’re Komsomols, the same people, the same faces. It’s the same show.”

David Ayerst quotes this correspondence in his book on The Guardian, and says it read “like a letter to end all communication”. So it did; I was finished with moderate men of all shades of opinion forever more.

Leaving Nazi Germany for neutral Switzerland, he says he had a pretty good idea even at the time how everything was going to end. And I believe him. By temperament, he expects everything to end in horror and madness and total collapse of civilization, so props to him for choosing the proper time and place for his temperament to be exactly correct. He writes:

All this likewise indubitably belonged to history, and would have to be historically assessed; like the Murder of the Innocents, or the Black Death, or the Battle of Paschendaele. But there was something else; a monumental death-wish, an immense destructive force loosed in the world which was going to sweep over everything and everyone, laying them flat, burning, killing, obliterating, until nothing was left. Those German agronomes in their green uniform suits with feathers in their hats – they had their part to play. So had the paunchy Brown-Shirts, and the matronly blonde maidens painting swastikas on the windows of Jewish shops. So had the credulous armies of the just, listening open-mouthed to Intourist patter, or seeking reassurance from a boozy sandalled Wicksteed. Wise old Shaw, high-minded old Barbusse, the venerable Webbs, Gide the pure in heart and Picasso the impure, down to poor little teachers, crazed clergymen and millionaires, drivelling dons and very special correspondents like Duranty, all resolved, come what might, to believe anything, however preposterous, to overlook anything, however villainous, to approve anything, however obscurantist and brutally authoritarian, in order to be able to preserve intact the confident expectation that one of the most thorough-going, ruthless, and bloody tyrannies ever to exist on Earth could be relied on to champion human freedom, the brotherhood of man, and all the other good liberal causes to which they had dedicated their lives. All resolved, in other words, to abolish themselves and their world, the rest of us with it. Nor have I from that time ever had the faintest expectation that, in earthly terms, anything could be salvaged; that any earthly battle could be won or earthly solution found. It has all just been sleep-walking to the end of the night.

II.

Muggeridge’s description of World War II is actually super hilarious.

I was not expecting this. When you take one of the darkest and most pessimistic writers of the twentieth century and put him in the middle of one of the twentieth century’s greatest horrors, you might expect the result to have at least a touch of grimness about it, or at least not to leave you rolling on the floor laughing. You would be wrong.

Muggeridge, inspired by some force even he did not understand, decided to enlist in the British military when the war broke out. He’s a bit too old by this point to be a front-line infantryman, and his intellect, connections, and experience with foreign countries catch the eye of Military Intelligence. They recruit him as a spy. His first job is counter-intelligence – hanging around in the army, making sure that there aren’t any secret German spies there. Well, there either aren’t any secret German spies, or else they’re at least not saying that they’re secret German spies, so this task turns out to be kind of a combination of boring, useless, and hilarious. He describes a typical day:

I find it difficult to recall what regular duties I had, if any…Our section was supposed to be responsible for securing the Headquarters from the incursions of enemy agents who might pry out its secrets or subvert its personnel. This gave us a free hand to do almost anything and go almost anywhere. If we went drinking in pubs, it was to keep a look-out for suspicious characters; if we pikced up girls, it was to probe their intentions in frequenting the locality.

A fellow-officer told me of how, on a pub-crawl, ostensibly a security reconnaissance, he got drunk, and, as was his way when in such a condition, pretended to be a foreigner, using strange gestures and speaking with an accent. The next day, badly hung over, he was sent a report of the movements of a suspicious foreigner, and told to check up on them. Tracing the suspect’s movements from pub to pub, it slowly dawned on him he was following himself the night before. When he told me of his adventure, to comfort him I said that it was what we were all doing all the time – keeping ourselves under close surveillance. This was what security was all about.

In a similar vein, another FS officer, idly thumbing over the Security List – a top-secret document containing the names of all subjects who were to be at once apprehended if they tried to get into or out of the country – found he was in it.

Graham Greene was a very famous early 20th century author. Like pretty much every other famous early 20th century author, he was a good friend of Malcolm Muggeridge’s. Greene was working in another branch of Intelligence at the time, and they needed someone for a secret mission, and Greene mooted Muggeridge’s name. He found himself plucked out of his cushy job drinking at pubs and tracking himself, and sent to MI6’s secret spy school at Bletchley Park, where he was taught various hilariously impractical skills like how to make invisible ink out of bird poop. He was then sent on a secret mission to Mozambique, so that just in case anything relevant to World War II were to happen in Mozambique, Her Majesty’s Government would have a secret agent in place.

The Mozambique chapters were among the funniest of the entire book. The Germans and Italians, inspired by the same principle, had also sent agents to Mozambique. It was not at all hard to figure out who they were, nor was Muggeridge’s identity particularly hard to figure out. There was only one nice hotel in Mozambique, so Muggeridge, the German spy, and the Italian spy all got rooms there and spent most of the time glaring at each other during communal dinners, or lying on the beach an appropriate distance away from one another, keeping watch.

Sometimes they would engage in hilarious secret plots against each other. Muggeridge, after chancing into a friendship with a member of Mozambique’s small German community, arranged for his friend to tell the German spy that he was only faking friendship with Muggeridge so he could steal his secrets for the good of The Reich. He then proceeded to “rob” Muggeridge’s house (with Muggeridge’s gleeful consent), producing for his German “master” a trove of documents which, when decoded, suggested that the Italian spy was secretly working for the British. This caused a big fight between the German spy and the Italian spy, which given that there wasn’t really much to spy on in Mozambique, was considered a fantastic success for the British cause and raised Muggeridge’s standing as some kind of intelligence prodigy.

Later in the war, Mozambique actually became sort of relevant as troop convoys started sailing by. Muggeridge bribed local officials to keep a watch out, and ended up foiling a very real German plot to do some sort of vague thing involving ships – as a result, when the war started winding down to the point where maintaining a presence in Mozambique was no longer viewed as entirely necessary, he came home and was promoted into the inner circles of intelligence. His new position was under Kim Philby, the head of the Department Of Counter-Intelligence Against The Soviet Union, who turned out to be a really bad choice for this position given that he, LIKE EVERY OTHER PROGRESSIVE INTELLECTUAL IN THE ENTIRE COUNTRY OF BRITAIN, was a secret Soviet spy. But at the time he seemed okay enough, and he sent Muggeridge to France to aid in the Liberation there.

We like to think of the Liberation of France as a nice, happy time, but for Muggeridge it was basically the time when an entire country worth of very angry Frenchmen massacred, pogrommed, lynched, or otherwise descended upon anyone accused of collaborating with the German occupation. Unsurprisingly, everybody turned out to think their personal and political rivals had collaborated with the German occupation, so it was basically the atmosphere of a 17th century Massachusetts witch hunt, only with less restraint.

Muggeridge’s job was, as usual, darkly hilarious – actual spies for the French and British governments usually acted all cooperative toward the German occupation to keep their cover and get a chance of infiltrating enemy ranks; as a result, they were usually First Up Against The Wall When The Liberation Came. Sure, they said “I was just a spy doing it as part of a secret plan,” but of course everybody said that. So Muggeridge had to rush from prison to prison, trying to convince mobs of angry Frenchmen not to execute the people who had just been most instrumental in saving them.

His spy career ended with what seems like maybe the most typical incident in the entire book – somehow P. G. Wodehouse had wandered into Nazi Germany and been stuck in a prison camp there. Then he had wandered out into France, gotten marked as a Collaborator, and was now in serious fear for his life. The British Secret Service picked Muggeridge as their Official Attache For P. G. Wodehouse Related Affairs, showing such exceptional genius in choosing the right man for the job that you would think they would have been able to get AT LEAST ONE ANTI-SOVIET COUNTERINTELLIGENCE AGENT WHO WASN’T A SECRET SOVIET SPY. Anyway, Muggeridge and Wodehouse wander around the cratered, mob-ruled French landscape, having a series of very Wodehousian adventures, until finally the war ends, Wodehouse is deposited safely the United States, and Muggeridge is able to return to Britain.

The book ends with the funeral of Sidney Webb, the early socialist hero his family idolized, who died just after World War II. Muggeridge is invited to the event because his wife is a distant cousin of the Webb family; he has to hold his nose throughout. At the time of his death, Webb is more beloved than ever by a grateful populace. His and his wife’s great works, Soviet Communism: A New Civilisation and The Truth About Soviet Russia, have become Bibles of the Left and part of Stalin’s cult of personality. Their opponents, the sorts who say that maybe Stalin isn’t the reincarnation of Christ, have been summarily dispatched – Muggeridge describes one of his friends from the journalism world, a reporter universally respected for helping expose Nazi atrocities, who made the mistake of trying to do the same with Soviet atrocities:

When Voigt turned the furious indignation with which he had lambasted the Nazi terror on to Stalin’s, his former liberal friends and associates discovered in him a Nazi sympathizer. Another liberal newspaper, the News Chronicle, ran an article about [his publication] headlined HITLER’S FAVORITE READING, with pictures of the Fuhrer and Voigt looking amicably across at one another.

In other words, Webb dies at the height of his career, his lies unexposed. George Bernard Shaw writes a letter to the newspapers suggesting that a man of Webb’s standing deserves a national hero’s funeral, everyone agrees, and he and his wife are interred in Westminster Abbey before a crowd of dignitaries including the Prime Minister (despite their own atheism and specific demands not to be placed in a church).

Muggeridge watches the whole sordid spectacle – the Dean of the Cathedral singing the praises of an unrepentant atheist “whose crowning achievement had been to commend to his fellow-countrymen and the whole world as a new civilization a system of servitude more far-reaching and comprehensive than any hitherto known” and ends his book very abruptly, saying only that “Another way has to be found and explored.”

III.

And then he dies before writing any more volumes of his autobiography, let alone telling us what the other way is.

He quotes very approvingly, as the heart of his philosophy, a passage by his friend Hugh Kingsmill:

What is divine in man is elusive and impalpable, and he is easily tempted to embody it in a concrete form – a church, a country, a social system, a leader – so that he may realize it with less effort and serve it with more profit. Yet the attempt to externalize the kingdom of heaven in a temporal shape must end in disaster. It cannot be created by charters or constitutions nor established by arms. Those who seek for it alone will reach it together, and those who seek it in company will perish by themselves.

And indeed, he writes a lot about how the whole problem started when people started being utopian and getting it into their heads to fix things on earth, rather than seek for “treasure in heaven”.

Some atheists I know write a lot about how religious people think you should hate the world because it’s awful and only some future world, ie Heaven, can be any good. Some religious people I know write a lot about how that’s total poppycock. Certainly G. K. Chesterton would have said something about how the world being sinful and full of flaws is not a reason to hate it, but precisely why we should love it, and Leah Libresco would say something about how hating the world is Gnosticism and Gnosticism is a heresy.

But I think Muggeridge might be pretty close to the atheist straw man on this point, with the key exception that religion isn’t what made him hate the world. He started off hating the world, and religion and mysticism offered him something not to hate, some way to say “Okay, but there’s some divinity buried in all this mess”. He is brilliant, he is compassionate, he is a great writer, it’s impossible to read his autobiography without loving him – but that he hates the world is hard to deny. I write sometimes about how beliefs that we consider abominable can sometimes be therapeutic mental crutches for people with the right cast of mind, and Muggeridge certainly found the idea of the world as a vale of suffering that would soon melt away to be oddly comforting in times of distress.

On the other hand, I’m not sure what to make of his opposition to trying to fix things here on Earth. He clearly hated Stalinism. When he hated Stalinism, he reacted by trying to make there be less Stalinism, which seems like a very reasonable thing to do. But the Communists hated capitalism. They reacted by trying to make there be less capitalism. Other than Muggeridge being right about the object-level issue and the Communists being wrong, it’s hard to see what the difference in principle is between them. The best I can do – and I worry I’m doing great violence to his intellectual uniqueness by rounding him off to my own ways of thinking – is to view him as suggesting some sort of precautionary principle, like that before you make a change you should be sure it’s something that has worked before (like non-Stalinism) and not a totally new idea (like Stalinism). But I am pretty sure if I suggested that to him he would roll his eyes and tell me that I’m such a modern and I don’t get it at all.

@slatestarcodex   
"But if we stop dumping raw sewage into…"  
"YOU'RE SO UTOPIAN!"  
"I just thoug-"  
"STOP TRYING TO IMMANENTIZE THE ESCHATON!"

— Scott Alexander (@slatestarcodex) June 25, 2013

@slatestarcodex One day we shall bring forth a new species of man free from utopian desires and cleansed of belief in human perfectibility.

— Scott Alexander (@slatestarcodex) June 25, 2013

The one thing I can be really sure of is that Muggeridge doesn’t want us to get stuck again in the same position we were in during the 30s and 40s where we totally ignored Stalin’s crimes due to our own political biases. Okay. I respect that. It was really eye-opening seeing exactly how brainwashed the entire European, British, and American Left were, and the whole situation gave me a lot more understanding of how overwhelmingly the Question of Communism dominated intellectual and political life in the first half of the century.

I was born in the 80s, at the very tail end of the Cold War, when we’d all had the decency to put all the Communists in one country and all the capitalists in another and make them express their differences like civilized men – ie by pointing thousands of hair-trigger nuclear missiles at one another. In the early days of Communism, we just didn’t know. Would Russia go Communist? Would Germany? Would France? Would everywhere? Muggeridge talks about how one of Britain’s main concerns in post-Liberation France was that the entire country would just move en masse to Communism as soon as the Nazis were out, which somehow or other mysteriously failed to happen EVEN THOUGH EVERY SINGLE ONE OF THE WESTERN AGENTS SENT TO PREVENT THAT WAS SECRETLY WORKING FOR THE SOVIETS.

And then the Cold War started, and this very gradually settled down to an equilibrium where okay, a lot of the Western intelligentsia stayed Communist, but at least they had the decency to realize that it was unpopular and the Revolution probably wasn’t literally going to happen next week.

By coincidence, just last week I read about the sad death of historian Robert Conquest, the man who was able to succeed where Muggeridge failed and drag Britain and America kicking and screaming into admitting Stalin wasn’t such a great guy. Conquest had one great advantage over Muggeridge, which was that he wrote in 1968 when, far from being our allies in a world war, the Soviets were technically our Cold War enemies and we were sort of okay with hearing bad things about them. But even then, he faced an extraordinary uphill battle. The most famous legend about him involved the second edition of his book, which came out right around the time the Soviet Union fell and its indisputable records of Stalin’s famines and purges became public knowledge. He supposedly asked to have the new version titled I TOLD YOU SO, YOU FUCKING FOOLS.

This part of our intellectual history is kind of forgotten. Who hears about Sidney and Beatrice Webb nowadays? Who hears about Walter Duranty? Yet these people during their times were absolute titans, “thought leaders” in the modern terminology – as per Muggeridge, Duranty “came to be accepted as the great Russian expert in America, and played a major part in shaping President Roosevelt’s policies vis-a-vis the USSR”. We hear a lot about our moral failures in terms of not stopping the Holocaust, but our quarter-century complicity with and even adulation of Stalinism seems like one of those facts that just fell by the wayside.

A lot of people think that I’m too easy on crackpots, or too fond of contrarians, or too interested in protecting witches, or whatever. But hearing all of these stories about the universal progressive Western adulation of Stalin is really scary. It’s way too easy for the darkest and most primal parts of my brain to map neatly onto the modern modalities of rejecting and punishing disagreement. “Really? You think this random journalist who isn’t even a trained Kremlinologist knows more than expert consensus?” “Killing millions of people, oh God, you’re one of those conspiracy losers.” “It’s obvious you’re just a privileged white guy who’s already decided to believe anything that reflects negatively on Slavs and foreigners.” “Although we respect free speech, that doesn’t extend to pro-Nazi propaganda and worker’s-paradise denialism.” Part of my respect for contrarians is that contrarianism is this incredibly fragile and precious art which needs to be kept alive for the times it is needed – rare times, times that hopefully won’t come up in our lifetimes, but times that, when they do come, desperately need a core of people willing to stand up to the establishment. Cultivating contrarianism is a lot like owning a gun – you get a heck of a lot of opportunities to shoot yourself in the foot, but also very occasionally one opportunity to save your life.

But then, on the other hand, here’s Muggeridge again:

Solzhenitsyn has provided the perfect parable on this theme with his description of Mrs. Roosevelt’s conducted visit to a labor camp where he was doing time. The estimable lady, who spawned the moral platitudes of the contemporary liberal wisdom as effortlessly and plenteously as the most prolific salmon, was easily persuaded that the camp in question was a humanely conducted institution for curing the criminally inclined. A truly wicked woman would have been ashamed to be so callous and so gullible.

Really? Gullible how? I’m sure the Soviets were moderately competent in making sure Roosevelt didn’t see anything too untoward. So what was she supposed to do?

I think of those people who say the US government is setting up FEMA interment camps as we speak to imprison dissenters against the New World Order. They provide some things that look sort of like evidence – photos (which turn out to be of random prisons or, in one case, an Amtrak station), documents (which turn out to be out of context references to setting up FEMA refugee camps for people displaced by disasters), et cetera. The people talking about this are total loons.

But Type 1 errors trade off against Type 2 errors. Make absolutely sure you’re the sort of person who never misses a Stalinist gulag, and you become the type of person who’s easy prey for the FEMA internment camp theory. Make absolutely sure you don’t believe in FEMA internment camps, and you’re liable to miss a Stalinist gulag as soon as the Soviet government gets Duranty to print “Oh, don’t worry, that’s just an Amtrak station”. Use the heuristic of “just trust expert consensus, experts always know what they’re talking about”, and you are now one of the tens of thousands of grateful readers who helped make Sidney and Beatrice Webb’s Soviet Communism: A New Civilisation into a best-seller.

What I’m saying is – there is no royal road. This is why I think learning rationality and the art of sifting through evidence is so important.

As for Muggeridge? I’m not sure he has much to teach there. Yes, he deserves the thanks of a grateful civilization for being a lone voice in the wilderness warning us about Stalin. But after that, as per his Wikipedia page, he was a lone voice in the wilderness warning us about contraception. After that, he became a lone voice in the wilderness warning us about marijuana. After that, he became a lone voice in the wilderness warning us about blasphemy in The Life Of Brian.

I am glad there are all types of people in the world. I am glad that there are crotchety, contrarian, cynical old reporters who constantly feel like everything is hurling off the precipice into Hell, because when things are actually hurling off the precipice into Hell, these people are the first to notice. In the same way, I am glad that there are dedicated survivalists who stockpile canned food in underground shelters in case of the nuclear apocalypse, because if there is ever an actual nuclear apocalypse, these people will survive and rebuild the human race.

But I am not digging a bomb shelter myself, and I am pretty sure I cannot bring myself to be quite as cynical as Malcolm Muggeridge.

# Stop Adding Zeroes

Dylan Matthews writes a critique of effective altruism. There is much to challenge in it, and some has already been challenged by people like Ryan Carey. Perhaps I will go into it at more length later. But for now I want to discuss a specific argument of Matthews’. He writes – and I am editing liberally to keep it short, so be sure to read the whole thing:

Nick Bostrom — the Oxford philosopher who popularized the concept of existential risk — estimates that about 10^54 human life-years (or 10^52 lives of 100 years each) could be in our future if we both master travel between solar systems and figure out how to emulate human brains in computers.

Even if we give this 10^54 estimate “a mere 1% chance of being correct,” Bostrom writes, “we find that the expected value of reducing existential risk by a mere one billionth of one billionth of one percentage point is worth a hundred billion times as much as a billion human lives.”

Put another way: The number of future humans who will never exist if humans go extinct is so great that reducing the risk of extinction by 0.00000000000000001 percent can be expected to save 100 billion more lives than, say, preventing the genocide of 1 billion people. That argues, in the judgment of Bostrom and others, for prioritizing efforts to prevent human extinction above other endeavors. This is what X-risk obsessives mean when they claim ending world poverty would be a “rounding error.”

[…]

These arguments give a false sense of statistical precision by slapping probability values on beliefs. But those probability values are literally just made up. Maybe giving $1,000 to the Machine Intelligence Research Institute will reduce the probability of AI killing us all by 0.00000000000000001. Or maybe it’ll make it only cut the odds by 0.00000000000000000000000000000000000000000000000000000000000000001. If the latter’s true, it’s not a smart donation; if you multiply the odds by 10^52, you’ve saved an expected 0.0000000000001 lives, which is pretty miserable. But if the former’s true, it’s a brilliant donation, and you’ve saved an expected 100,000,000,000,000,000,000,000,000,000,000,000 lives.

I don’t have any faith that we understand these risks with enough precision to tell if an AI risk charity can cut our odds of doom by 0.00000000000000001 or by only 0.00000000000000000000000000000000000000000000000000000000000000001. And yet for the argument to work, you need to be able to make those kinds of distinctions.

Matthews correctly notes that this argument – often called “Pascal’s Wager” or “Pascal’s Mugging” – is on very shaky philosophical ground. The AI risk movement generally agrees, and neither depends on it nor uses it very often. Nevertheless, this is what Matthews wants to discuss. So let’s discuss it.

His argument is that sure, it looks like fighting existential risk and saving 10^54 people is important. But that depends exactly how small the chance of your anti-x-risk plan working is. He gives two different possibilities which, if you count the zeroes, turn out to be 10^-17 and 10^-66. Then he asks: which one is it, 10^-17 or 10^-66? We just don’t know.

Well, actually, we do know. It’s probably not the 10^-66 one, because nothing is ever 10^-66 and you should never use that number.

Let me try to justify this.

Consider which of the following seems intuitively more likely:

First, that a well-meaning person donates $1000 to MIRI or FLI or FHI, this aids their research and lobbying efforts, and as a result they are successfully able to avert an unfriendly superintelligence.

Or second, that despite our best efforts, a research institute completes an unfriendly superintelligence. They are seconds away from running the program for the first time when, just as the lead researcher’s finger hovers over the ENTER key, a tornado roars into the laboratory. The researcher is sucked high into the air. There he is struck by a meteorite hurtling through the upper atmosphere, which knocks him onto the rooftop of a nearby building. He survives the landing, but unfortunately at precisely that moment the building is blown up by Al Qaeda. His charred corpse is flung into the street nearby. As the rubble settles, his face is covered by a stray sheet of newspaper; the headline reads 2016 PRESIDENTIAL ELECTION ENDS WITH TRUMP AND SANDERS IN PERFECT TIE. In small print near the bottom it also lists the winning Powerball numbers, which perfectly match those on a lottery ticket in the researcher’s pocket. Which is actually kind of funny, because he just won the same lottery last week.

Well, the per-second probability of getting sucked into the air by a tornado is 10^-12; that of being struck by a meteorite 10^-16; that of being blown up by a terrorist 10^-15. The chance of the next election being Sanders vs. Trump is 10^-4, and the chance of an election ending in an electoral tie about 10^-2. The chance of winning the Powerball is 10^-8 so winning it twice in a row is 10^-16. Chain all of those together, and you get 10^-65. On the other hand, Matthews thinks it’s perfectly reasonable to throw out numbers like 10^-66 when talking about the effect of x-risk donations. To take that number seriously is to assert that the second scenario is ten times more likely than the first!

In Made Up Statistics, I discuss how sometimes our system one intuitive reasoning and system two mathematical reasoning can act as useful checks on each other. A commenter described this as “sometimes it’s better to pull numbers out of your ass and use them to get an answer, than to pull an answer out of your ass.”

A good example of this is 80,000 Hours’ page on why people shouldn’t get too excited about medicine as an altruistic career (oops). They argue that the good a doctor does by treating illnesses is minimal compared to the good she can do by earning to give. Their reasoning goes like this: the average doctor saves 4 QALYs a year through medical interventions. The average doctor’s salary is $150,000 or so; if she donates 10% to charity, that’s $15,000. As per Givewell, that kind of money could save 300 QALYs per year. The value of the earning to give is so much higher then the value of the actual doctoring that you might as well skip the doctoring entirely and go into whatever earns you the most money.

Intuitively, people’s system 1s think “Doctor? That’s something where you’re saving lots of lives, so it must be a good altruistic career choice.” But then when you pull numbers out of your ass, it turns out not to be. Crucially, exactly which numbers you pull out of your ass doesn’t matter much as long as they’re remotely believable. 80,000 Hours tried their best to figure out how many QALYs doctors save per year, but this is obviously a really difficult question and for all we know they could be off by an order of magnitude. The point is, it doesn’t matter. They could be off by a figure of ten times, twenty times, even fifty times and it wouldn’t affect their argument. I’ve gone over their numbers with them and it’s really, really, really hard to remotely believably make the “number of QALYs saved per doctor” figure come out high enough to challenge the earning-to-give route. Sure, you’re pulling numbers out of your ass, but even your ass has some standards.

It’s the same with Matthews’ estimates about x-risk. He intuitively thinks that x-risk charities can’t be that great compared to fighting global poverty or whatever other good cause. He (very virtuously) decides to double-check that assumption with numbers, even if he has to make up the numbers himself. The problem is, he doesn’t have a very good feel for numbers of that size, so he thinks he can literally make up whatever numbers he wants, instead of doing something that we jokingly call “making up whatever number you want” but which in fact involves some sanity checks to make sure they’re remotely believable proxies for our intuitions. He thinks “I don’t expect x-risk charities to work very well, so what the heck, I might as well call that 10^-66”, whereas he should be thinking something like “10^-66 means about ten times less likely than my chance of getting tornado-meteor-terrorist-double-lottery-Trumped in any particular second, is that a remotely believable approximation to how unlikely I think existential risk is?”

Just as it is very hard to come up with a remotely believable number that spoils 80,000 Hours’ anti-doctor argument, so you have to really really stretch your own credulity to come up with numbers where Bostrom’s x-risk argument doesn’t work.

(some people argue that LW-style rationality is a bad idea, because you can’t really think with probabilities. I would argue that even if that’s true, there is at least a small role for rationality in avoiding being bamboozled by other people trying to think with probabilities and doing it wrong. This is a modest claim, but no more modest than Wittgenstein’s view of philosophy, which was that it was a useful thing to know in order to protect yourself from taking philosophers too seriously.)

But one more point. Suppose Matthews’ intuition is indeed that the chance of AI risk charities working out is precisely ten times less than his per-second chance of getting tornado-meteor-terrorist-double-lottery-Trumped. In that case, I offer him the following whatever-the-opposite-of-a-gift is: we can predict pretty precisely the yearly chance of a giant asteroid hitting our planet, it’s way more than 10^-66, and the whole x-risk argument applies to it just as well as to AI or anything else. What now?

Because this isn’t just about defending the particular proposition of AI. It’s a more general principle of staring into the darkness. If you try to be good, if you don’t let yourself fiddle with your statistical intuitions until they give you the results you want, sometimes you end up with weird or scary results.

Like that a person who wants to cure as much disease as possible would be better off becoming a hedge fund manager than a doctor.

Or that your charity dollar would be better sent off to sub-Saharan Africa to purchase something called “praziquantel” than given to the sad-looking man with the cardboard sign you see on the way to work.

Or that a person who wants to reduce suffering in the world should focus almost obsessively on chickens.

One of the founding beliefs of effective altruism is that when math tells you something weird, you at least consider trusting the math. If you’re allowed to just add on as many zeroes as it takes to justify your original intuition, you miss out on the entire movement.

Everyone has their own idea of what trusting the math entails and how far they want to go with it. Some people go further than I do. Other people go less far. But anybody who makes a good-faith effort to trust it even a little is, in my opinion, an acceptable ally worth including in the effective altruist tent. They have abandoned a nice safe chance to donate to the local symphony and feel good about themselves, in favor of a life of feeling constantly uncomfortable with their decisions, looking extremely silly to normal people, and having Dylan Matthews write articles in Vox calling them “white male autistic nerds”.

Matthews is firmly underneath the effective-altruist tent. He writes that he’s worried that a focus on existential risk will detract from the causes he really cares about, like animal rights. He gets very, very excited about animal rights, and in his work for Vox he’s done some incredible work promoting them. Good! I also donate to animal rights’ charities and I think we need far more people who do that.

And yet, the same arguments he deploys against existential risk could be leveled against him also – “how can you worry about chickens when there are millions of families trying to get by on minimum wage? Effective altruists need to stop talking about animals if they ever want to attract anybody besides white males into the movement.” What then?

Malcolm Muggeridge describes a vision he once had, of everyone in the world riding together on a giant train toward realms unknown. Each person wants to get off at their own stop, but when the train comes to their station, the engineer speeds right by. All the other passengers laugh and hoot and sing the praises of the engineer, because this means the train will get to their own stations faster. But of course each one finds that when the train comes to their station, why, it speeds past that one too, and they are left to rage impotently at the unfairness.

And I worry that Matthews is urging us to shoot past the “existential risk” station in order to get to the “animal rights” station a little faster, without reflecting on the likely consequences.

This certainly isn’t to say we all need to get off at the first station. I myself am very interested in existential risk, but I give less than a third of my donations to x-risk related charities (no, I can’t justify this, it’s a sanity-preserving exception). I respect those who give more. I also respect those who give less. Existential risk isn’t the most useful public face for effective altruism – everyone incuding Eliezer Yudkowsky agrees about that. But at least allowing people interested in x-risk into the tent and treating them respectfully seems like an inescapable consequence of the focus on reason and calculation that started effective altruism in the first place.

# Figure/Ground Illusions

There’s a social justice concept called “distress of the privileged”. It means that if some privileged group is used to having things 100% their own way, and then some reform means that they only get things 99% their own way, this feels from the inside like oppression, like the system is biased against them, like now the other groups have it 100% their own way and they have it 0% and they can’t understand why everyone else is being so unfair.

I’ve said before that I think a lot of these sorts of ideas are poor fits for the one-sided issues they’re generally applied to, but more often accurate in describing the smaller, more heavily contested ideological issues where most of the explicit disputes lie nowadays. And so there’s an equivalent to distress of the privileged where supporters of a popular ideology think anything that’s equally fair to popular and unpopular ideologies, or even biased toward the popular ideology less than everyone else, is a 100%-against-them super-partisan tool of the unpopular people.

So I want to go back to Dylan Matthews’ article about EA. He is concerned that there’s too much focus on existential risk in the movement, writing:

Effective altruism is becoming more and more about funding computer science research to forestall an artificial intelligence–provoked apocalypse.

And:

EA Global was dominated by talk of existential risks, or X-risks.

And:

What was most concerning was the vehemence with which AI worriers asserted the cause’s priority over other cause areas.

And:

The movement has a very real demographic problem, which contributes to very real intellectual blinders of the kind that give rise to the AI obsession.

It sounds like he worries AI concerns are taking over the movement, that they’ve become the dominant strain, that all anybody’s interested in is AI.

Here is the latest effective altruist survey. This survey massively overestimates concern with AI risks, because only the AI risk sites did a good job publicizing the survey. Nevertheless, it still finds that of 813 effective altruists, only 77 donated to the main AI risk charity listed, the Machine Intelligence Research Institute. In comparison, 211 – almost three times as many – donated to the Against Malaria Foundation (note that not all participants donated to any cause, and some may have donated to several)

An explicit question about areas of concern tells a similar story – out of ten multiple-choice areas of concern, AI risks, x-risks, and the far future are 5th, 7th, and last respectively. The top is, once again, global poverty.

I wasn’t at the EA Summit and can’t talk about it from a position of personal knowledge. But the program suggests that out of thirty or so different events, just one was explicitly about AI, and two others were more generically x-risk related. The numbers at the other two EA summits were even less impressive. In Melbourne, there was only one item related to AI or x-risk – putting it on equal footing with the “Christianity And Effective Altruism” talk.

I do hear that the Bay Area AI event got special billing, but I think this was less because only AI is important, and more because some awesome people like Elon Musk were speaking, whereas a lot of the other panels featured people so non-famous that they even very briefly flirted with trying to involve me.

And – when people say that you should donate all of your money to AI risk and none to any other cause, they may well be thinking in terms of a world where about $50 billion is donated to global poverty yearly, and by my estimates the total budget for AI risk is less than $5 million a year. There are world-spanning NGOs like UNICEF and the World Bank working on global poverty and employing tens of thousands of people; in contrast, I bet > 10% of living AI risk researchers have been to one of Alicorn’s weekly dinner parties, and her table is only big enough for six people at a time. In this context, on the margin, “you should make your donation to AI” means “I think AI should get more than 1/10,000th of the pot”.

I suspect that “AI is dominating the effective altruist movement”, when you look at it, means “AI is given an equal place at the effective altruist table, compared to being totally marginalized everywhere else.” By figure-ground illusion, that makes it seem “dominant”.

Or consider me personally. I probably sound like some kind of huge AI partisan by this point, but I give less than a third of my donations to AI related causes, and if you ask me whether you should donate to them, I will tell you that I honestly don’t know. The only reason I keep speaking out in favor of AI risks is that when everyone else is so sure about it, my “I don’t know” suddenly becomes a far-fringe position that requires defending more than less controversial things. By figure-ground illusion, that makes me seem super-pro-AI.

In much the same way, I have gotten many complaints that the comments section of this blog leans way way way to the right, whereas the survey (WHICH I WILL ONE DAY POST, HONEST) suggests that it is almost perfectly evenly balanced. I can’t prove that the median survey-taker is also the median commenter, but I think probably people used to discussions entirely dominated by the left are seeing an illusory conservative bias in a place where both sides are finally talking equally.

Less measurably, I think I get this with my own views: – I despair of ever shaking the label of “neoreactionary sympathizer” just for treating them with about the same level of respect and intellectual interest I treat everyone else. And I despair of ever shaking the label of “violently obsessively anti-social-justice guy” – despite a bunch of posts expressing cautious support for social justice causes – just because I’m not willing to give them a total free pass when they do something awful, or totally demonize their enemies, in the same way as the median person I see on Facebook.

Or at least this is how it feels from the inside. Maybe this is how everybody feels from the inside, and Ayatollah Khameini is sitting in Tehran saying “I am so confused by everything that I try to mostly maintain an intellectual neutrality in which I give Islam exactly equal time to every other religion, but everyone else is unfairly hostile to it so I concentrate on that one, and then people call me a fanatic.” It doesn’t seem likely. But I guess it’s possible.

# My Id On Defensiveness

I.

I’ll admit it – I’ve been unusually defensive lately. Defensive about Hallquist’s critique of rationalism, defensive about Matthews’ critique of effective altruism, and if you think that’s bad you should see my Tumblr.

Brienne noticed this and asked me why I was so defensive all the time, and I thought about it, and I realized that my id had a pretty good answer. I’m not sure I can fully endorse my id on this one, but it was a sufficiently complete and consistent picture that I thought it was worth laying out.

I like discussion, debate, and reasoned criticism. But a lot of arguments aren’t any of those things. They’re the style I describe as ethnic tension, where you try to associate something you don’t like with negative affect so that other people have an instinctive disgust reaction to it.

There are endless sources of negative affect you can use. You can accuse them of being “arrogant”, “fanatical”, “hateful”, “cultish” or “refusing to tolerate alternative opinions”. You can accuse them of condoning terrorism, or bullying, or violence, or rape. You can call them racist or sexist, you can call them neckbeards or fanboys. You can accuse them of being pseudoscientific denialist crackpots.

If you do this enough, the group gradually becomes disreputable. If you really do it enough, the group becomes so toxic that it becomes somewhere between a joke and a bogeyman. Their supporters will be banned on site from all decent online venues. News media will write hit pieces on them and refuse to ask for their side of the story because ‘we don’t want to give people like that a platform’. Their concerns will be turned into bingo cards for easy dismissal. People will make Facebook memes strawmanning them, and everyone will laugh in unison and say that yep, they’re totally like that. Anyone trying to correct the record will be met with an “Ew, gross, this place has gone so downhill that the [GROUP] is coming out of the woodwork!” and totally ignored.

(an easy way to get a gut feeling for this – go check how they talk about liberals in very conservative communities, then go check how they talk about conservatives in very liberal communities. I’m talking about groups that somehow manage to gain this status everywhere simultaneously)

People like to talk a lot about “dehumanizing” other people, and there’s some debate over exactly what that entails. Me, I’ve always thought of it the same was as Aristotle: man is the rational animal. To dehumanize them is to say their ideas don’t count, they can’t be reasoned with, they no longer have a place at the table of rational discussion. And in a whole lot of Internet arguments, doing that to a whole group of people seems to be the explicit goal.

II.

There’s a term in psychoanalysis, “projective identification”. It means accusing someone of being something, in a way that actually turns them into that thing. For example, if you keep accusing your (perfectly innocent) partner of always being angry and suspicious of you, eventually your partner’s going to get tired of this and become angry, and maybe suspicious that something is up.

Declaring a group toxic has much the same effect. The average group has everyone from well-connected reasonable establishment members to average Joes to horrifying loonies. Once the group starts losing prestige, it’s the establishment members who are the first to bail; they need to protect their establishment credentials, and being part of a toxic group no longer fits that bill. The average Joes are now isolated, holding an opinion with no support among experts and trend-setters, so they slowly become uncomfortable and flake away as well. Now there are just the horrifying loonies, who, freed from the stabilizing influence of the upper orders, are able to up their game and be even loonier and more horrifying. Whatever accusation was leveled against the group to begin with is now almost certainly true.

I have about a dozen real-world examples of this, but all of them would be so mind-killing as to dominate the comments to the exclusion of my actual point, so generate them on your own and then shut up about them – in the meantime, I will use a total hypothetical. So consider Christianity.

Christianity has people like Alvin Plantinga and Ross Douthat who are clearly very respectable and key it into the great status-conferring institutions like academia and journalism. It has a bunch of middle-class teachers and plumbers and officer workers who go to church and raise money to send Bibles to Africa and try not to sin too much. And it has horrifying loons who stand on street corners waving signs saying “GOD HATES FAGS” and screaming about fornicators.

Imagine that Christianity suffers a sudden total dramatic in prestige, to the point where wearing a cross becomes about as socially acceptable as waving a Confederate flag. The New York Times fires Ross Douthat, because they can’t tolerate people like that on their editorial staff. The next Alvin Plantinga chooses a field other than philosophy of religion, because no college would consider granting him tenure for that.

With no Christians in public life or academia, Christianity starts to seem like a weird belief that intelligent people never support, much like homeopathy or creationism. The Christians have lost their air support, so to speak. The average college-educated individual starts to feel really awkward about this, and they don’t necessarily have to formally change their mind and grovel for forgiveness, they can just – go to church a little less, start saying they admire Jesus but they’re not Christian Christian, and so on.

Gradually the field is ceded more and more to the people waving signs and screaming about fornicators. The opponents of Christianity ramp up their attacks that all Christians are ignorant and hateful, and this is now a pretty hard charge to defend against, given the demographic. The few remaining moderates, being viewed suspiciously in churches that are now primarily sign-waver dominated and being genuinely embarrassed to be associated with them, bail at an increased rate, leading their comrades to bail at an even faster rate, until eventually it is entirely the sign wavers.

Then everybody agrees that their campaign against Christians was justified all along, because look how horrible Christians are, they’re all just a bunch of sign-wavers who have literally no redeeming features. Now even if the original pressure that started the attack on Christianity goes away, it’s inconceivable that it will ever come back – who would join a group that is universally and correctly associated with horrible ignorant people?

(I think this is sort of related to what Eliezer calls evaporative cooling of group beliefs, but not quite the same.)

In quite a number of the most toxic and hated groups around, I feel like I can trace a history where the group once had some pretty good points and pretty good people, until they were destroyed from the outside by precisely this process.

In Part I, I say that sometimes groups can get so swamped by other people’s insults that they turn toxic. There’s nothing in Part I to suggest that this would be any more than a temporary setback. But because of this projective identification issue, I think it’s way more than that. It’s more like there’s an event horizon, a certain amount of insulting and defamation you can take after which you will just get more and more hated and your reputation will never recover.

III.

There is some good criticism, where people discuss the ways that groups are factually wrong or not very helpful, and then those groups debate that, and then maybe everyone is better off.

But the criticism that makes me defensive is the type of criticism that seems to be trying to load groups with negative affect in the hopes of pushing them into that event horizon so that they’ll be hated forever.

I support some groups that are a little weird, and therefore especially vulnerable to having people try to push them into the event horizon.

And as far as I can tell, the best way to let that happen is to let other people load those groups with negative affect and do nothing about it. The average person doesn’t care whether the negative affect is right or wrong. They just care how many times they see the group’s name in close proximity to words like “crackpot” or “cult”.

I judge people based on how likely they are to do this to me. One reason I’m so reluctant to engage with feminists is that I feel like they constantly have a superweapon pointed at my head. Yes, many of them are very nice people who will never use the superweapon, but many others look like very nice people right up to the point where I disagree with them in earnest at which point they vaporize me and my entire social group.

On the other hand, you can push people into the event horizon, but you can’t pull them in after you. That means that the safest debate partners, the ones you can most productively engage, will be the people who have already been dismissed by everyone else. This is why I find talking to people like ClarkHat and JayMan so rewarding. They are already closer to the black hole than I am, and so they have no power to load me with negative affect or destroy my reputation. This reduces them to the extraordinary last resort of debating with actual facts and evidence. Even better, it gives me a credible reason to believe that they will. Schelling talks about “the right to be sued” as an important right that businesses need to protect for themselves, not because anyone likes being sued, but because only businesses that can be sued if they slip up have enough credibility to attract customers. In the same way, there’s a “right to be vulnerable to attack” which is almost a necessary precondition of interesting discussion these days, because only when we’re confronted with similarly vulnerable people can we feel comfortable opening up.

IV.

But with everybody else? I don’t know.

I remember seeing a blog post by a moderately-well known scholar – I can’t remember who he was or find the link, so you’ll just have to take my word for it – complaining that some other scholar in the field who disagreed with him was trying to ruin his reputation. Scholar B was publishing all this stuff falsely accusing Scholar A of misconduct, calling him a liar and a fraud, personally harassing him, and falsely accusing Scholar A of personally harassing him (Scholar B). This kinda went back and forth between both scholars’ blogs, and Scholar A wrote this heart-breaking post I still (sort of) remember, where he notes that he now has a reputation in his field for “being into drama” and “obsessed with defending himself” just because half of his blog posts are arguments presenting evidence that Scholar B’s fraudulent accusations are, indeed fraudulent.

It is really easy for me to see the path where rationalists and effective altruists become a punch line and a punching bag. It starts with having a whole bunch of well-publicized widely shared posts calling them “crackpots” and “abusive” and “autistic white men” without anybody countering them, until finally we end up in about the same position as, say, Objectivism. Having all of those be wrong is no defense, unless somebody turns it into such. If no one makes it reputationally costly to lie, people will keep lying. The negative affect builds up more and more, and the people who always wanted to hate us anyway because we’re a little bit weird say “Oh, phew, we can hate them now”, and then I and all my friends get hated and dehumanized, the prestigious establishment people jump ship, and there’s no way to ever climb out of the pit. All you need for this to happen is one or two devoted detractors, and boy do we have them.

That seems to leave only two choices.

First, give up on ever having the support of important institutions like journalism and academia and business, slide into the black hole, and accept decent and interesting conversations with other black hole denizens as a consolation prize while also losing the chance at real influence or attracting people not already part of the movement.

Or, second, call out every single bad argument, make the insults and mistruths reputationally costly enough that people think at least a little before doing them – and end up with a reputation for being nitpicky, confrontational and fanatical all the time.

(or, as the old Tumblr saying goes, “STOP GETTING SO DEFENSIVE EVERY TIME I ATTACK YOU!”)

I don’t know any third solution. If somebody does, I would really like to hear it.

# The Goddess of Everything Else

[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 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 daughers 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.

# On Overconfidence

[Epistemic status: This is basic stuff to anyone who has read the Sequences, but since many readers here haven’t I hope it is not too annoying to regurgitate it. Also, ironically, I’m not actually that sure of my thesis, which I guess means I’m extra-sure of my thesis]

I.

A couple of days ago, the Global Priorities Project came out with a calculator that allowed you to fill in your own numbers to estimate how concerned you should be with AI risk. One question asked how likely you thought it was that there would be dangerous superintelligences within a century, offering a drop down menu with probabilities ranging from 90% to 0.01%. And so people objected: there should be options to put in only one a million chance of AI risk! One in a billion! One in a…

For example, a commenter writes that: “the best (worst) part: the probability of AI risk is selected from a drop down list where the lowest probability available is 0.01%!! Are you kidding me??” and then goes on to say his estimate of the probability of human-level (not superintelligent!) AI this century is “very very low, maybe 1 in a million or less”. Several people on Facebook and Tumblr say the same thing – 1/10,000 chance just doesn’t represent how sure they are that there’s no risk from AI, they want one in a million or more.

Last week, I mentioned that Dylan Matthews’ suggestion that maybe there was only 10^-67 chance you could affect AI risk was stupendously overconfident. I mentioned that was thousands of lower than than the chance, per second, of getting simultaneously hit by a tornado, meteor, and al-Qaeda bomb, while also winning the lottery twice in a row. Unless you’re comfortable with that level of improbability, you should stop using numbers like 10^-67.

But maybe it sounds like “one in a million” is much safer. That’s only 10^-6, after all, way below the tornado-meteor-terrorist-double-lottery range…

So let’s talk about overconfidence.

Nearly everyone is very very very overconfident. We know this from experiments where people answer true/false trivia questions, then are asked to state how confident they are in their answer. If people’s confidence was well-calibrated, someone who said they were 99% confident (ie only 1% chance they’re wrong) would get the question wrong only 1% of the time. In fact, people who say they are 99% confident get the question wrong about 20% of the time.

It gets worse. People who say there’s only a 1 in 100,000 chance they’re wrong? Wrong 15% of the time. One in a million? Wrong 5% of the time. They’re not just overconfident, they are fifty thousand times as confident as they should be.

This is not just a methodological issue. Test confidence in some other clever way, and you get the same picture. For example, one experiment asked people how many numbers there were in the Boston phone book. They were instructed to set a range, such that the true number would be in their range 98% of the time (ie they would only be wrong 2% of the time). In fact, they were wrong 40% of the time. Twenty times too confident! What do you want to bet that if they’d been asked for a range so wide there was only a one in a million chance they’d be wrong, at least five percent of them would have bungled it?

Yet some people think they can predict the future course of AI with one in a million accuracy!

Imagine if every time you said you were sure of something to the level of 999,999/1 million, and you were right, the Probability Gods gave you a dollar. Every time you said this and you were wrong, you lost $1 million (if you don’t have the cash on hand, the Probability Gods offer a generous payment plan at low interest). You might feel like getting some free cash for the parking meter by uttering statements like “The sun will rise in the east tomorrow” or “I won’t get hit by a meteorite” without much risk. But would you feel comfortable predicting the course of AI over the next century? What if you noticed that most other people only managed to win $20 before they slipped up? Remember, if you say even one false statement under such a deal, all of your true statements you’ve said over years and years of perfect accuracy won’t be worth the hole you’ve dug yourself.

Or – let me give you another intuition pump about how hard this is. Bayesian and frequentist statistics are pretty much the same thing [citation needed] – when I say “50% chance this coin will land heads”, that’s the same as saying “I expect it to land heads about one out of every two times.” By the same token, “There’s only a one in a million chance that I’m wrong about this” is the same as “I expect to be wrong on only one of a million statements like this that I make.”

What do a million statements look like? Suppose I can fit twenty-five statements onto the page of an average-sized book. I start writing my predictions about scientific and technological progress in the next century. “I predict there will not be superintelligent AI.” “I predict there will be no simple geoengineering fix for global warming.” “I predict no one will prove P = NP.” War and Peace, one of the longest books ever written, is about 1500 pages. After you write enough of these statements to fill a War and Peace sized book, you’ve made 37,500. You would need to write about 27 War and Peace sized books – enough to fill up a good-sized bookshelf – to have a million statements.

So, if you want to be confident to the level of one-in-a-million that there won’t be superintelligent AI next century, you need to believe that you can fill up 27 War and Peace sized books with similar predictions about the next hundred years of technological progress – and be wrong – at most – once!

This is especially difficult because claims that a certain form of technological progress will not occur have a very poor track record of success, even when uttered by the most knowledgeable domain experts. Consider how Nobel-Prize winning atomic scientist Ernest Rutherford dismissed the possibility of nuclear power as “the merest moonshine” less than a day before Szilard figured out how to produce such power. In 1901, Wilbur Wright told his brother Orville that “man would not fly for fifty years” – two years later, they flew, leading Wilbur to say that “ever since, I have distrusted myself and avoided all predictions”. Astronomer Joseph de Lalande told the French Academy that “it is impossible” to build a hot air balloon and “only a fool would expect such a thing to be realized”; the Montgolfier brothers flew less than a year later. This pattern has been so consistent throughout history that sci-fi titan Arthur C. Clarke (whose own predictions were often eerily accurate) made a heuristic out of it under the name Clarke’s First Law: “When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.”

Also – one good heuristic is to look at what experts in a field think. According to Muller and Bostrom (2014), a sample of the top 100 most-cited authors in AI ascribed a > 70% probability to AI within a century, a 50% chance of superintelligence conditional on human-level, and a 10% chance of existential catastrophe conditional on human level AI. Multiply it out, and you get a couple percent chance of superintelligence-related existential catastrophe in the next century.

Note that my commenter wasn’t disagreeing with the 4% chance. They were disagreeing with the possibility that there would be human-level AI at all, that is, the 70% chance! That means that he was saying, essentially, that he was confident he could write a million sentences – that is, twenty-seven War and Peace‘s worth – all of which were trying to predict trends in a notoriously difficult field, all of which contradicted a well-known heuristic about what kind of predictions you should never try to make, all of which contradicted the consensus opinion of the relevant experts – and only have one of the million be wrong!

But if you feel superior to that because you don’t believe there’s only a one-in-a-million chance of human-level AI, you just believe there’s a one-in-a-million chance of existential catastrophe, you are missing the point. Okay, you’re not 300,000 times as confident as the experts, you’re only 40,000 times as confident. Good job, here’s a sticker.

Seriously, when people talk about being able to defy the experts a million times in a notoriously tricky area they don’t know much about and only be wrong once – I don’t know what to think. Some people criticize Eliezer Yudkowsky for being overconfident in his favored interpretation of quantum mechanics, but he doesn’t even attach a number to that. For all I know, maybe he’s only 99% sure he’s right, or only 99.9%, or something. If you are absolutely outraged that he is claiming one-in-a-thousand certainty on something that doesn’t much matter, shouldn’t you be literally a thousand times more outraged when every day people are claiming one-in-a-million level certainty on something that matters very much? It is almost impossible for me to comprehend the mindsets of people who make a Federal Case out of the former, but are totally on board with the latter.

Everyone is overconfident. When people say one-in-a-million, they are wrong five percent of the time. And yet, people keep saying “There is only a one in a million chance I am wrong” on issues of making really complicated predictions about the future, where many top experts disagree with them, and where the road in front of them is littered with the bones of the people who made similar predictions before. HOW CAN YOU DO THAT?!

II.

I am of course eliding over an important issue. The experiments where people offering one-in-a-million chances were wrong 5% of the time were on true-false questions – those with only two possible answers. There are other situations where people can often say “one in a million” and be right. For example, I confidently predict that if you enter the lottery tomorrow, there’s less than a one in a million chance you will win.

On the other hand, I feel like I can justify that. You want me to write twenty-seven War and Peace volumes about it? Okay, here goes. “Aaron Aaronson of Alabama will not win the lottery. Absalom Abramowtiz of Alaska will not win the lottery. Achitophel Acemoglu of Arkansas will not win the lottery.” And so on through the names of a million lottery ticket holders.

I think this is what statisticians mean when they talk about “having a model”. Within the model where there are a hundred million ticket holders, and we know exactly one will be chosen, our predictions are on very firm ground, and our intuition pumps reflect that.

Another way to think of this is by analogy to dart throws. Suppose you have a target that is half red and half blue; you are aiming for red. You would have to be very very confident in your dart skills to say there is only a one in a million chance you will miss it. But if there is a target that is 999,999 millionths red, and 1 millionth blue, then you do not have to be at all good at darts to say confidently that there is only a one in a million chance you will miss the red area.

Suppose a Christian says “Jesus might be God. And he might not be God. 50-50 chance. So you would have to be incredibly overconfident to say you’re sure he isn’t.” The atheist might respond “The target is full of all of these zillions of hypotheses – Jesus is God, Allah is God, Ahura Mazda is God, Vishnu is God, a random guy we’ve never heard of is God. You are taking a tiny tiny submillimeter-sized fraction of a huge blue target, painting it red, and saying that because there are two regions of the target, a blue region and a red region, you have equal chance of hitting either.” Eliezer Yudkowsky calls this “privileging the hypothesis”.

There’s a tougher case. Suppose the Christian says “Okay, I’m not sure about Jesus. But either there is a Hell, or there isn’t. Fifty fifty. Right?”

I think the argument against this is that there are way more ways for there not to be Hell than there are for there to be Hell. If you take a bunch of atoms and shake them up, they usually end up as not-Hell, in much the same way as the creationists’ fabled tornado-going-through-a-junkyard usually ends up as not-a-Boeing-747. For there to be Hell you have to have some kind of mechanism for judging good vs. evil – which is a small part of the space of all mechanisms, let alone the space of all things – some mechanism for diverting the souls of the evil to a specific place, which same, some mechanism for punishing them – again same – et cetera. Most universes won’t have Hell unless you go through a lot of work to put one there. Therefore, Hell existing is only a very tiny part of the target. Making this argument correctly would require an in-depth explanation of formalizations of Occam’s Razor, which is outside the scope of this essay but which you can find on the LW Sequences.

But this kind of argumentation is really hard. Suppose I predict “Only one in 150 million chance Hillary Clinton will be elected President next year. After all, there are about 150 million Americans eligible for the Presidency. It could be any one of them. Therefore, Hillary covers only a tiny part of the target.” Obviously this is wrong, but it’s harder to explain how. I would say that your dart-aim is guided by an argument based on a concrete numerical model – something like “She is ahead in the polls by X right now, and candidates who are ahead in the polls by X usually win about 50% of the time, therefore, her real probability is more like 50%.”

Or suppose I predict “Only one in a million chance that Pythagoras’ Theorem will be proven wrong next year.” Can I get away with that? I can’t quite appeal to “it’s been proven”, because there might have been a mistake in (all the) proofs. But I could say: suppose there are five thousand great mathematical theorems that have undergone something like the level of scrutiny as Pythagoras’, and they’ve been known on average for two hundred years each. None of them have ever been disproven. That’s a numerical argument that the rate of theorem-disproving is less than one per million years, and I think it holds.

Another way to do this might be “there are three hundred proofs of Pythagoras’ theorem, so even accepting an absurdly high 10%-per-proof chance of being wrong, the chance is now only 10^-300.” Or “If there’s a 10% chance each mathematician reading a proof missing something, and one million mathematicians have read the proof of Pythagoras’ Theorem, then the probability that they all missed it is more like 10^-1,000,000.”

But this can get tricky. Suppose I argued “There’s a good chance Pythagoras’ Theorem will be disproven, because of all Pythagoras’ beliefs – reincarnation, eating beans being super-evil, ability to magically inscribe things on the moon – most have since been disproven. Therefore, the chance of a randomly selected Pythagoras-innovation being wrong is > 50%.”

Or: “In 50 past presidential elections, none have been won by women. But Hillary Clinton is a woman. Therefore, the chance of her winning this election is less than 1/50.”

All of this stuff about adjusting for size of the target or for having good mathematical models is really hard and easy to do wrong. And then you have to add another question: are you sure, to a level of one-in-a-million, that you didn’t mess up your choice of model at all?

Let’s bring this back to AI. Suppose that, given the complexity of the problem, you predict with utter certainty that we will not be able to invent an AI this century. But if the modal genome trick pushed by people like Greg Cochran works out, within a few decades we might be able to genetically engineer humans far smarter than any who have ever lived. Given tens of thousands of such supergeniuses, might we be able to solve an otherwise impossible problem? I don’t know. But if there’s a 1% chance that we can perform such engineering, and a 1% chance that such supergeniuses can invent artificial intelligence within a century, then the probability of AI within the next century isn’t one in a million, it’s one in ten thousand.

Or: consider the theory that all the hard work of brain design has been done by the time you have a rat brain, and after that it’s mostly just a matter of scaling up. You can find my argument for the position in this post – search for “the hard part is evolving so much as a tiny rat brain”. Suppose there’s a 10% chance this theory is true, and a 10% chance that researchers can at least make rat-level AI this century. Then the chance of human-level AI is not one in a million, but one in a hundred.

Maybe you disagree with both of these claims. The question is: did you even think about them before you gave your one in a million estimate? How many other things are there that you never thought about? Now your estimate has, somewhat bizarrely, committed you to saying there’s a less than one in a million chance we will significantly enhance human intelligence over the next century, and a less than one in a million chance that the basic-scale-up model of intelligence is true. You may never have thought directly about these problems, but by saying “one in a million chance of AI in the next hundred years”, you are not only committing yourself to a position on them, but committing yourself to a position with one-in-a-million level certainty even though several domain experts who have studied these fields for their entire lives disagree with you!

A claim like “one in a million chance of X” not only implies that your model is strong enough to spit out those kinds of numbers, but that there’s only a one in a million chance you’re using the wrong model, or missing something, or screwing up the calculations.

A few years ago, a group of investment bankers came up with a model for predicting the market, and used it to design a trading strategy which they said would meet certain parameters. In fact, they said that there was only a one in 10^135 chance it would fail to meet those parameters during a given year. A human just uttered the probability “1 in 10^135”, so you can probably guess what happened. The very next year was the 2007 financial crisis, the model wasn’t prepared to deal with the extraordinary fallout, the strategy didn’t meet its parameters, and the investment bank got clobbered.

This is why I don’t like it when people say we shouldn’t talk about AI risk because it involves “Knightian uncertainty”. In the real world, Knightian uncertainty collapses back down to plain old regular uncertainty. When you are an investment bank, the money you lose because of normal uncertainty and the money you lose because of Knightian uncertainty are denominated in the same dollars. Knightian uncertainty becomes just another reason not to be overconfident.

III.

I came back to AI risk there, but this isn’t just about AI risk.

You might have read Scott Aaronson’s recent post about Aumann Agreement Theorem, which says that rational agents should be able to agree with one another. This is a nice utopian idea in principle, but in practice, well, nobody seems to be very good at carrying it out.

I’d like to propose a more modest version of Aumann’s agreement theorem, call it Aumann’s Less-Than-Total-Disagreement Theorem, which says that two rational agents shouldn’t both end up with 99.9…% confidence on opposite sides of the same problem.

The “proof” is pretty similar to the original. Suppose you are 99.9% confident about something, and learn your equally educated, intelligent, and clear-thinking friend is 99.9% confident of the opposite. Arguing with each other and comparing your evidence fails to make either of you budge, and neither of you can marshal the weight of a bunch of experts saying you’re right and the other guy is wrong. Shouldn’t the fact that your friend, using a cognitive engine about as powerful as your own, got so heavily different a conclusion make you worry that you’re missing something?

But practically everyone is walking around holding 99.9…% probabilities on the opposite sides of important issues! I checked the Less Wrong Survey, which is as good a source as any for people’s confidence levels on various tough questions. Of the 1400 respondents, about 80 were at least 99.9% certain that there were intelligent aliens elsewhere in our galaxy; about 170 others were at least 99.9% certain that they weren’t. At least 80 people just said they were certain to one part in a thousand and then got the answer wrong! And some of the responses were things like “this box cannot fit as many zeroes as it would take to say how certain I am”. Aside from stock traders who are about to go bankrupt, who says that sort of thing??!

And speaking of aliens, imagine if an alien learned about this particular human quirk. I can see them thinking “Yikes, what kind of a civilization would you get with a species who routinely go around believing opposite things, always with 99.99…% probability?”

Well, funny you should ask.

I write a lot about free speech, tolerance of dissenting ideas, open-mindedness, et cetera. You know which posts I’m talking about. There are a lot of reasons to support such a policy. But one of the big ones is – who the heck would burn heretics if they thought there was a 5% chance the heretic was right and they were wrong? Who would demand that dissenting opinions be banned, if they were only about 90% sure of their  
 own? Who would start shrieking about “human garbage” on Twitter when they fully expected that in some sizeable percent of cases, they would end up being wrong and the garbage right?

Noah Smith recently asked why it was useful to study history. I think at least one reason is to medicate your own overconfidence. I’m not just talking about things like “would Stalin have really killed all those people if he had considered that he was wrong about communism” – especially since I don’t think Stalin worked that way. I’m talking about Neville Chamberlain predicting “peace in our time”, or the centuries when Thomas Aquinas’ philosophy was the preeminent Official Explanation Of Everything. I’m talking about Joseph “no one will ever build a working hot air balloon” Lalande. And yes, I’m talking about what Muggeridge writes about, millions of intelligent people thinking that Soviet Communism was great, and ending out disastrously wrong. Until you see how often people just like you have been wrong in the past, it’s hard to understand how uncertain you should be that you are right in the present. If I had lived in 1920s Britain, I probably would have been a Communist. What does that imply about how much I should trust my beliefs today?

There’s a saying that “the majority is always wrong”. Taken literally it’s absurd – the majority thinks the sky is blue, the majority don’t believe in the Illuminati, et cetera. But what it might mean, is that in a world where everyone is overconfident, the majority will always be wrong about which direction to move the probability distribution in. That is, if an ideal reasoner would ascribe 80% probability to the popular theory and 20% to the unpopular theory, perhaps most real people say 99% popular, 1% unpopular. In that case, if the popular people are urging you to believe the popular theory more, and the unpopular people are urging you to believe the unpopular theory more, the unpopular people are giving you better advice. This would create a strange situation in which good reasoners are usually engaged in disagreeing with the majority, and also usually “arguing for the wrong side” (if you’re not good at thinking probablistically, and almost no one is), but remain good reasoners and the ones with beliefs most likely to produce good outcomes. Unless you count “why are all of our good reasoners being burned as witches?” as a bad outcome.

I started off by saying this blog was about “the principle of charity”, but I had trouble defining it and in retrospect I’m not that good at it anyway. What can be salvaged from such a concept? I would say “behave the way you would if you were less than insanely overconfident about most of your beliefs.” This is the Way. The rest is just commentary.

Discussion Questions (followed by my own answers in ROT13)

1. What is your probability that there is a god? (Svir creprag)  
2. What is your probability that psychic powers exist? (Bar va bar gubhfnaq)  
3. What is your probability that anthropogenic global warming will increase temperatures by at least 1C by 2050? (Avargl creprag)  
4. What is your probability that a pandemic kills at least one billion people in a 5 year period by 2100? (Svsgrra creprag)  
5. What is your probability that humans land on Mars by 2050? (Rvtugl creprag)  
6. What is your probability that superintelligent AI (=AI better than almost every human at almost every cognitive task) exists by 2115? (Gjragl svir creprag)

# Probabilities Without Models

[Epistemic status: Not original to me. Also, I might be getting it wrong.]

A lot of responses to my Friday post on overconfidence centered around this idea that we shouldn’t, we can’t, use probability at all in the absence of a well-defined model. The best we can do is say that we don’t know and have no way to find out. I don’t buy this:

“Mr. President, NASA has sent me to warn you that a saucer-shaped craft about twenty meters in diameter has just crossed the orbit of the moon. It’s expected to touch down in the western United States within twenty-four hours. What should we do?”

“How should I know? I have no model of possible outcomes.”

“Should we put the military on alert?”

“Maybe. Maybe not. Putting the military on alert might help. Or it might hurt. We have literally no way of knowing.”

“Maybe we should send a team of linguists and scientists to the presumptive landing site?”

“What part of ‘no model’ do you not understand? Alien first contact is such an inherently unpredictable enterprise that even speculating about whether linguists should be present is pretending to a certainty which we do not and cannot possess.”

“Mr. President, I’ve got our Israeli allies on the phone. They say they’re going to shoot a missile at the craft because ‘it freaks them out’. Should I tell them to hold off?”

“No. We have no way of predicting whether firing a missile is a good or bad idea. We just don’t know.”

In real life, the President would, despite the situation being totally novel and without any plausible statistical model, probably make some decision or another, like “yes, put the military on alert”. And this implies a probability judgment. The reason the President will put the military on alert, but not, say, put banana plantations on alert, is that in his opinion the aliens are more likely to attack than to ask for bananas.

Fine, say the doubters, but surely the sorts of probability judgments we make without models are only the most coarse-grained ones, along the lines of “some reasonable chance aliens will attack, no reasonable chance they will want bananas.” Where “reasonable chance” can mean anything from 1% to 99%, and “no reasonable chance” means something less than that.

But consider another situation: imagine you are a director of the National Science Foundation (or a venture capitalist, or an effective altruist) evaluating two proposals that both want the same grant. Proposal A is by a group with a long history of moderate competence who think they can improve the efficiency of solar panels by a few percent; their plan is a straightforward application of existing technology and almost guaranteed to work and create a billion dollars in value. Proposal B is by a group of starry-eyed idealists who seem very smart but have no proven track record; they say they have an idea for a revolutionary new kind of super-efficient desalinization technology; if it works it will completely solve the world’s water crisis and produce a trillion dollars in value. Your organization is risk-neutral to a totally implausible degree. What do you do?

Well, it seems to me that you choose Proposal B if you think it has at least a 1/1000 chance of working out; otherwise, you choose Proposal A. But this requires at least attempting to estimate probabilities in the neighborhood of 1/1000 without a model. Crucially, there’s no way to avoid this. If you shrug and take Proposal A because you don’t feel like you can assess proposal B adequately, that’s making a choice. If you shrug and take Proposal B because what the hell, that’s also making a choice. If you are so angry at being placed in this situation that you refuse to choose either A or B and so pass up both a billion and a trillion dollars, that’s a choice too. Just a stupid one.

Nor can you cry “Pascal’s Mugging!” in order to escape the situation. I think this defense is overused and underspecified, but at the very least, it doesn’t seem like it can apply in places where the improbable option is likely to come up over your own lifespan. So: imagine that your organization actually reviews about a hundred of these proposals a year. In fact, it’s competing with a bunch of other organizations that also review a hundred or so such proposals a year, and whoever’s projects make the most money gains lots of status and new funding. Now it’s totally plausible that, over the course of ten years, it might be a better strategy to invest in things that have a one in a thousand chance of working out. Indeed, maybe you can see the organizations that do this outperforming the organizations that don’t. The question really does come down to your judgment: are Project B’s odds of success greater or less than 1/1000?

Nor is this a crazy hypothetical situation. A bunch of the questions we have to deal with come down to these kinds of decisions made without models. Like – should I invest for retirement, even though the world might be destroyed by the time I retire? Should I support the Libertarian candidate for president, even though there’s never been a libertarian-run society before and I can’t know how it will turn out? Should I start learning Chinese because China will rule the world over the next century? These questions are no easier to model than ones about cryonics or AI, but they’re questions we all face.

The last thing the doubters might say is “Fine, we have to face questions that can be treated as questions of probability. But we should avoid treating them as questions of probability anyway. Instead of asking ourselves ‘is the probability that the desalinization project will work greater or less than 1/1000’, we should ask ‘do I feel good about investing this money in the desalinization plant?’ and trust our gut feelings.”

There is some truth to this. My medical school thesis was on the probabilistic judgments of doctors, and they’re pretty bad. Doctors are just extraordinarily overconfident in their own diagnoses; a study by Bushyhead, who despite his name is not a squirrel, found that when doctors were 80% certain that patients had pneumonia, only 20% would turn out to have the disease. On the other hand, the doctors still did the right thing in most every case, operating off of algorithms and heuristics that never mentioned probability. The conclusion was that as long as you don’t force doctors to think about about what they’re doing in mathematical terms, everything goes fine – something I’ve brought up before in the context of the Bayes mammogram problem. Maybe this generalizes. Maybe people are terrible at coming up with probabilities for things like investing in desalinization plants, but will generally make the right choice.

But refusing to frame choices in terms of probabilities also takes away a lot of your options. If you use probabilities, you can check your accuracy – the foundation director might notice that of a thousand projects she had estimated as having 1/1000 probabilities, actually about 20 succeeded, meaning that she’s overconfident. You can do other things. You can compare people’s success rates. You can do arithmetic on them (“if both these projects have 1/1000 probability, what is the chance they both succeed simultaneously?”), you can open prediction markets about them.

Most important, you can notice and challenge overconfidence when it happens. I said last post that when people say there’s only a one in a million chance of something like AI risk, they are being stupendously overconfident. If people just very quietly act as if there’s a one in a million chance of such risk, without ever saying it, then no one will ever be able to call them on it.

I don’t want to say I’m completely attached to using probability here in exatly the normal way. But all of the alternatives I’ve heard fall apart when you’ve got to make an actual real-world choice, like sending the military out to deal with the aliens or not.

[EDIT: Why regressing to meta-probabilities just gives you more reasons to worry about overconfidence]

[EDIT-2: “I don’t know”]

[EDIT-3: A lot of debate over what does or doesn’t count as a “model” in this case. Some people seem to be using a weak definition like “any knowledge whatsoever about the process involved”. Others seem to want a strong definition like “enough understanding to place this event within a context of similar past events such that a numerical probability can be easily extracted by math alone, like the model where each flip of a two-sided coin has a 50% chance of landing heads”. Without wanting to get into this, suffice it to say that any definition in which the questions above have “models” is one where AI risk also has a model.]

# Mysticism and Pattern-Matching

[Epistemic status: Total conjecture.]

One of the things that got me interested in psychiatry was the sheer weirdness of the human brain’s failure modes. We all hear that the brain is like a computer, but when a computer breaks, the screen goes black or it freezes or something. It doesn’t hear voices telling it that it’s Jesus, or start seeing tiny men running around on the floor. But for some reason, when the the human brain breaks, it may do exactly that. Why?

Psychiatry classes never just tell you the answer to this question, but reading between the lines I think it has something to do with top-down processing and pattern matching.



Bottom-up processing is when you go from basic elements to more complex ideas – for example, when you see the three letters C, A, and T in a row, you might combine them to get the the word CAT. Top-down processing is when more complex ideas change the way you interpret basic elements. For example, in the first picture above, the middle letters in both words are the same. We read the first as H, because the image as a gestalt suggests the word “THE” and the word “THE” suggests an H in the middle. We read the second as A, because the image as a gestalt suggests the word “CAT” and the word “CAT” has an A in the middle. Our big-picture idea has changed the way we view the smaller elements composing it.

The same is true of the second image. We recognize the phrase “PARIS IN THE SPRINGTIME”, and so we assume that’s what the sign is trying to show us. In fact, the sign doubles the word “the”. But since this is bizarre and not something that makes sense in the gestalt, we assume this is a mistake and gloss right over it. We do this very, very easily – how many times have I duplicated the word “the” in this essay already?

The third image is related to this tendency. To most people, it looks formless. Even once they hear that it’s an old black-and-white photograph of a cow’s head, it’s might still require a bit of staring before you catch on. But once you see the cow, the cow is obvious. It becomes impossible to see it as formless, impossible to see it as anything else. Having given yourself a top-down pattern to work from, the pattern automatically organizes the visual stimuli and makes sense of them.

This provides a possible explanation for hallucinations. Think of top-down processing as taking noise and organizing it to fit a pattern. Normally, you’ll only fit it to the patterns that are actually there. But if your pattern-matching system is broken, you’ll fit it to patterns that aren’t in the data at all.

The best example of this is Google Deep Dream:



I don’t know much about neural networks, so I may not be getting this entirely right, but as far as I understand it, they trained a neural network on some stimulus like a dog. This was for research in machine vision; they wanted the net to be able to recognize dogs when it saw them; to pattern-match potentially noisy images of dogs into its Platonic ideal of a dog. But if you turn the pattern-matching up, it will just start seeing dogs everywhere there’s even the slightest amount of noise that resembles a dog at all. You only matched the sign above to “PARIS IN THE SPRINGTIME” because it was almost exactly like that phrase; if we stick your pattern-matching software into overdrive, maybe every sentence would start looking like more meaningful alternatives. Eevn sceeentns wtih aolsmt all the lerttes rergaearnd wulod naelry ianslntty sanp itno pacle. Turn it all the way up, and maybe you could make every sentence look like “PARIS IN THE SPRINGTIME”. Or something.

So hallucinations are when your top-down processing/pattern-matching ability becomes so dysfunctional that it can generate people and objects out of random visual noise. Why it chooses some people and objects over others I don’t know, but it’s hardly surprising – it does the same thing every night in your dreams.

Many of the same people who have hallucinations also have paranoia. Paranoia seems to me to be overfunctioning of social pattern-matching. When Deep Dream sees the tiniest hint of a line here, a slight dark spot there, it pattern-matches it into an entire dog. When a paranoiac hears a stray word here, or sees a sideways glance there, they turn it into this vast social edifice of connected plots. Every new thing that happens is fit effortlessly into the same pattern. When their psychiatrist says they’re crazy, that gets fit into the pattern too – maybe the psychiatrist is a tool of the conspiracy, trying to confuse them into compliance.

So where does the mysticism come in?

I notice that the same people who have hallucinations also have mystical experiences. By mystical experiences, I don’t just mean “they see angels” – in that case, the relationship to hallucination would be a tautology. I mean they feel a sense of sudden understanding of and connection with the universe. I know at least three groups that do this: druggies, meditators, and prophets. The druggies report feelings of total understanding on their drugs, and also report hallucinations. The meditators occasionally achieve enlightenment, but look at any text about meditation and you find mentions of visions and hallucinations experienced during the practice. The voices heard by the prophets are too obvious to mention.

One well-known way of bringing on such experiences is to abuse your pattern-matching faculty. The Chicken Qabalah of Rabbi Lamed Ben Clifford (not really recommended) manages to link a pretty boring Bible verse to the letter yud, the creativity of God, the essence of existence, the sun, the phallus, the plane of Malkuth, and the number 496, then explains:

Like a mountain goat leaping ecstatically from crag to crag, one thought springs into another, and another, ad infinitum. You can continue, almost forever, connecting things that you never thought were connected. Sooner or later something’s going to snap and you will overcome the fundamental defect in your powers of perception.

And:

Was that the message Ezekiel was trying to convey? Probably not. But who cares! Whatever it was the old boy was originally trying to say shrinks to insignificance. It is far more important to my spiritual enlightenment that my mind was forced to churn at breakneck speed to put all of this together, and then open itself up to the infinite possibilities of meaning. Look hard enough at anything and eventually you will see everything! it doesn’t even have to make very much sense what you connect to what. It’s all ultimately connected!

This philosophy, which I associate both with kabbalah and with the more modern Western hermetic tradition, says that learning a set of extremely complicated correspondences is an important step toward gaining enlightenment. See for example this site, which helpfully relates the sephirah Netzach to the planet Venus, the number 7, the emerald, the lynx, the rose, cannabis, arsenic, copper, fire, the solar plexus chakra, the archangel Haniel, the Egyptian goddess Hathor, the concepts of love and victory, et cetera, et cetera. You’re supposed to be able to use this to interpret things – for example, if you have a dream about a lynx, it could correspond to anything else in the system – but it looks like it would quickly get unwieldy. And other sources will give completely different systems of correspondences, and nobody gets too upset over it – in fact, some sources will happily encourage you to come up with your own correspondences instead, as long as you stick to them. It seems like the goal is less “remember that it’s extremely important that emeralds correspond to lynxes in reality” and more “have some system, any system, of interesting correspondences in mind that you can apply to everything you come across”.

Nor does it especially matter what you’re interpreting. The traditional things to interpret are mysterious things like dreams, or the Bible, but Crowley famously performs a mystical analysis of Mother Goose nursery rhymes (see Interlude here). The important factor seems to be less about there being sacred truth in the object being analyzed, and more about the process of performing the analysis.

(Zen koans are a little different, but also sort of involve torturing a pattern-finding ability for apparently no reason)

So to skip to the point: I think all of this is about strengthening the pattern-matching faculty. You’re exercising it uselessly but impressively, the same way as the body-builder who lifts the same weight a thousand times until their arms are the size of tree trunks. Once the pattern-matching faculty is way way way overactive, it (spuriously) hallucinates a top-down abstract pattern in the whole universe. This is the experience that mystics describe as “everything is connected” or “all is one”, or “everything makes sense” or “everything in the universe is good and there for a purpose”. The discovery of a beautiful all-encompassing pattern in the universe is understandably associated with “seeing God”.

Religious scholar William James once experimented with nitrous oxide and reached a state where he felt he had total comprehension of the universe. According to a story which I can’t verify, he became infuriated at losing the thread of understanding once the chemical wore off, so he decided to take notes during the experience: write down the secrets of the universe then, and reread them once he was sober. The experiment completed, he picked up the notepad in feverish excitement, only to find that he had written OVERALL THERE IS A SMELL OF FRIED ONIONS.

Imagine one of those Google robots pointing at an empty patch of sky and saying “No, look, seriously, there’s a dog right there. Right there! How are you not seeing this?” Things that make perfect sense in the context of a state of overactive pattern-matching look meaningless to a pattern-matching faculty operating normally. At best, you can sort of see the lines of what seemed so clear before (“Yeah, I can see that that stain on the wall is vaguely dog-shaped.”) This matches the stories I’ve heard of people who have some mystical experience but then can’t maintain or recapture it.

I think other methods of inducing weird states of consciousness, like drugs and meditation, probably do the same thing by some roundabout route. Meditation seems like reducing stimuli, which is known to lead to hallucinations in eg sensory deprivation tanks or solitary confinement cells in jail. I think the general principle is that a low level of external stimuli makes your brain adjust its threshold for stimulus detection up until anything including random noise satisfies the threshold. As for drugs, there’s lots of reasons to think that the neurotransmission changes they create will alter the brain’s pattern processing strategies.

Things this hypothesis doesn’t explain: why mystical experiences are linked with a feeling of no time, no space, and no self; why prayer or extreme devotion seems to induce them (eg bhakti yoga), and why they can be so beneficial – that is, why do people with mystical experiences become happier and better adjusted? Maybe the feeling of the world making sense is naturally a pleasant and helpful one. Certainly the opposite can be very stressful!

# Magic Markers

[Thanks to some people on Rationalist Tumblr, especially prophecyformula, for help and suggestions.]

There’s an old philosophers’ saying – trust those who seek the truth, distrust those who say they’ve found it. The psychiatry version of this goes “Trust those who seek biological underpinnings for mental illness, distrust those who say they’ve found them.”

Niculescu et al (2015) say they’ve found them. Their paper describes a process by which they hunted for biomarkers – in this case changes in gene expression – that predict suicide risk among psychiatric patients. They test various groups of psychiatric patients (including post-mortem tissue from suicide victims) to find some plausible genes. Then they use those genes to predict suicidality in two cohorts of about 100 patients each, including people with depression, schizophrenia, schizoaffective disorder, and bipolar disorder. They arrive at an impressive 92% AUC – that being the area under the curve graphing sensitivity vs. specificity, a common measure of the accuracy with which they can distinguish people who will vs. won’t be suicidal in the future.

The science press, showing the skepticism and restraint for which they are famous, jump on board immediately. A New Blood Test Can Predict Whether A Patient Will Have Suicidal Thoughts With More Than 90% Accuracy, says Popular Science. New Blood Test Predicts Future Suicide Attempts, says PBS.

There is a procedure for this sort of thing. The procedure is that the rest of us sit back and quietly wait for James Coyne, author of How To Critique Claims For A Blood Test For Depression, to tell us exactly why it is wrong. But it’s been over a week now and this hasn’t happened and I’m starting to worry he’s asleep on the job. So even though this is somewhat outside my area of expertise, let me discuss a couple of factors that concern me about this study.

The 92% accuracy claim is for the authors’ model, called UP-SUICIDE, which combines 11 biomarkers and two clinical prediction instruments. A clinical prediction instrument is a test which asks questions like “How depressed are you feeling right now?” or “How many times have you attempted suicide before?”. By combining the predictive power of the eleven genes and two instruments, they managed to reach the 92% number advertised in the abstract.

It might occur to you to ask “Wait, a test in which you can just ask people if they’re depressed and hate their life sounds a lot easier than this biomarker thing. Are we sure that they’re not just getting all of their predictive power from there?”

The answer is: no, we’re not sure at all, and as far as I can tell the study goes to great pains in order to make it hard to tell to what degree they are doing this.

Conventional wisdom says that clinical instruments for predicting suicidality can attain AUCs of 0.74 to 0.88. This is most of the way to the 0.92 shown in the current study, but not quite as high. But the current study combines two different clinical prediction instruments. In Combining Scales To Assess Suicide Risk, a Spanish team combines a few different clinical prediction instruments to get an AUC of…0.92.

If you look really closely at Niculescu et al’s big results table, you find that each of the individual prediction instruments they use does almost as well – and in some cases better than – their UP-SUICIDE model as a whole. For example, when predicting suicidal ideation in all patients, the CFI-S instrument has an AUC of 0.89, compared to the entire model’s 0.92. When predicting suicide-related hospitalizations in depressed patients, the CFI-S has an AUC of 0.78, compared to the entire model’s 0.7. Here the biomarkers are just adding noise!

Are the cases where the entire model outperforms the CFI-S cases where the biomarkers genuinely help? We have no way of knowing. There are two clinical prediction instruments, the CFI-S and the SASS. Combined, they should outperform either one alone. So, for example, on suicidal ideation among all patients, the SASS has an AUC of 0.85, the CFI-S has an AUC of 0.89, and the model as a whole (both instruments combined + 11 biomarkers) has an AUC of 0.92. If we just combined the CFI-S and SASS, and threw out the biomarkers, would we do better or worse than 0.92? I don’t know and they don’t tell us. When all we’re doing is looking at the overall model, the biomarkers may be helping, hurting, or totally irrelevant.

So what if we throw out the clinical prediction instruments and just look at the biomarkers?

The authors use their panel of biomarkers for four different conditions: depression, bipolar, schizophrenia, and schizoaffective. And they have two different outcomes: suicidal ideation according to a test of such, and actual hospitalization for suicide. That’s a total of 4 x 2 = 8 tests that they’re conducting.

Of these eight different tests, the panel of biomarkers taken together come back insignificant on seven of them.

And there’s such a thing as “trending towards significance”, but this isn’t it. Here, I’ll give p-values:

Depression/ideation: p = 0.26  
Depression/hospitalization: p = 0.48  
Schizoaffective/ideation: p = 0.46  
Schizoaffective/hospitalization: p = 0.94  
Schizophrenia/ideation: p = 0.16  
Schizophrenia/hospitalization: p = 0.72  
Bipolar/hospitalization: p = 0.24

The only test of the eight that comes out significant is bipolar/ideation, where p = 0.007. This is fine (well, it’s fine if it’s supposed to be post-Bonferroni correction, which I can’t be sure of from the paper). But I notice three things. Number one, there were only 29 people in this group. Number two, some of the most impressive looking genes for the ideation condition were worthless for the hospitalization condition. CLIP4, which got p = 0.005 for the ideation condition, got p = 0.91 for the second condition and actually had negative predictive value. And third, some of the genes that best predicted bipolar in the validation data had no predictive value for bipolar at all in the training data, and were included only because they predicted major depressive disorder alone. Given that the effects jump across diagnoses and fail to carry over into even a slightly different method of assessing suicidality, this looks a lot less like a real finding and a lot more like a statistical blip.

Finally, note that even in bipolar ideation, their one apparent success, the biomarkers only got an AUC of 0.75, lower than either clinical predictive instrument. The only reason their model did better was because it added on the clinical predictive instruments themselves.

So here it looks like seven out of their eight tests failed miserably, one of them succeeded in a very suspicious way, and they covered over this by combining the data with the clinical predictive instruments which always worked very well. Then everyone interpreted this as the sexy and exciting result “biomarkers work!” rather than the boring result “biomarkers fail, but if you use other stuff instead you’ll still be okay.”

The absolute strongest conclusion you can draw from this study is “biomarkers may predict risk of suicidal ideation in bipolar disorder with an AUC of 0.75”. Instead, everyone thinks biomarkers predict suicidality and hospitalization in a set of four different disorders with AUC of 0.92, which is way beyond what the evidence can support.

II.

So much for that. Now let me explain why it wouldn’t matter much even if they were right.

AUC is a combination of two statistics called sensitivity and specificity. It’s a little complicated, but if we assume it means sensitivity and specificity are both 92% we won’t be far off.

Sensitivity is the probability that a randomly chosen positive case in fact tests positive. In this case, it means the probability that, if someone is actually going to be suicidal, the model flags them as high suicide risk.

Specificity is the probability that a randomly chosen negative case in fact tests negative. In this case, it means the probability that, if someone is not going to be suicidal, the model flags them as low suicide risk.

In this study population, about 7.5% of their patients are hospitalized for suicidality each year. So suppose you got a million depressed people similar to these. 75,000 would attempt suicide that year, and 925,000 wouldn’t.

Now, suppose you gave your million depressed people this test with a 92% sensitivity and specificity.

Of the 925,000 non-suicidal people, 92% – 851,000 – will be correctly evaluated as non-suicidal. 74,000- 8% – will be mistakenly evaluated as suicidal.

Of the 75,000 suicidal people, 92% – 69,000 – will be correctly evaluated as suicidal. 8% – 6,000 – will be mistakenly evaluated as non-suicidal.

But this means that of the 143,000 people the test says are suicidal, only 69,000 – less than half – actually will be!

So when people say “We have a blood test to diagnose suicidality with 92% accuracy!”, even if it’s true, what they mean is that they have a blood test which, if it comes back positive, there’s still less than 50-50 odds the person involved is suicidal. Okay. Say you’re a psychiatrist. There’s a 48% chance your patient is going to be suicidal in the next year. What are you going to do? Commit her to the hospital? I sure hope not. Ask her some questions, make sure she’s doing okay, watch her kind of closely? You’re a psychiatrist and she’s your depressed patient, you would have been doing that anyway. This blood test is not really actionable.

And then remember that this isn’t the blood test we have. We have some clinical prediction instruments that do this, and we have a blood test which maybe, if you are very trusting, diagnoses suicidality in bipolar disorder with 75% accuracy. At 75% sensitivity and specificity, only twenty percent of the people who test positive will be suicidal. So what?

There will never be a blood test for suicide that works 100%, because suicide isn’t 100% in the blood. I am the most biodeterminist person you know (unless you know JayMan), I am happy to agree with Martin and Tesser that that the heritability of learning Latin is 26% and the heritability of jazz is 45% and so on, but suicide is not just biological. Maybe people need some kind of biological predisposition to consider suicide. But whether they go ahead with it or not depends on whether they have a good or bad day, whether their partner breaks up with them, whether a friend hands them a beer and they get really drunk, et cetera. Taking all of this into account, it’s really unlikely that a blood test will ever get sensitive and specific enough to overcome these hurdles.

We should continue research on the biological underpinnings of depression and suicide, both for the sake of knowledge and because it might lead to better treatments. But having “a blood test for suicide” won’t be very useful, even if it works.

# Theses On Trump

The smart money still says Trump will crash and burn before getting the nomination. But everyone saying this should have to add “however, he’s certainly lasted much longer than we originally predicted.”

Scott Adams places the blame for this surprising perseverance on Trump-specific factors – namely, his brilliance as a persuader and manipulator. The mainstream media reports I’ve seen place it on Republican-specific factors, some combination of the words “ignorant”, “bigoted”, “white”, and “aggrieved entitlement”, followed by “therefore, Trump” – without a good explanation for why not therefore Scott Walker or therefore Rick Santorum.

I would argue it reflects a more general trend. As Exhibit A, I bring before the court Jeremy Corbyn.

In case you’re not following the United Kingdom: Labour, the UK’s major leftist party, is having leadership elections. Jeremy Corbyn is expected to win. In most ways, Corbyn is the diametric opposite of Trump. Trump is on the right, Corbyn is on the far left. Trump is inconsistent in his policy stances to the degree he even has any; Corbyn has strong opinions which he never deviates from even an inch. Trump has never held office; Corbyn has toiled in an unimpressive Parliamentary back bench position for decades. Trump is famously loud and bombastic, Corbyn is famously quiet and reserved, and likes to talk about things like gardening and how more people should be beekeepers.

But their rises to power look weirdly similar. Corbyn spent twenty years on the edges of British politics, mostly a figure of fun for his weird out-of-touch positions. When he announced the intention to run for Labour leader, everyone thought of it as a quixotic attempt to gain some free coverage, like when Dennis Kucinich runs for President. Some of the MPs who signed off on his candidacy request forms said they did it because they figured it wouldn’t matter one way or the other so they might as well make him happy. The media talked about how silly it was that he was even running at all. Then he started gathering momentum. The establishment freaked out and told everyone he was unacceptable and they were not to vote for him under any circumstances. A parade of important figures and personalities went on the news to personally beg voters not to vote for him under any circumstances, ably parodied by the Twitter account @corbynwarnings:

As Prime Minister, Corbyn Would Renationalise Children, Warns MP

— Corbyn Warnings (@CorbynWarnings) August 11, 2015

Corbyn 'Obsessed With Destroying The Moon', Warns MP

— Corbyn Warnings (@CorbynWarnings) August 11, 2015

Antibiotics Not Effective Against Corbyn, Warn Doctors

— Corbyn Warnings (@CorbynWarnings) August 13, 2015

Estimated 100% Of People Who Vote For Corbyn Will Die Within A Century, Warn Scientists

— Corbyn Warnings (@CorbynWarnings) August 12, 2015

Nevertheless, Corbyn seems amply placed to win the election, and British bookmakers give 3:1 odds in his favor. What happened?

Well, for one thing, this is the first year people are allowed to vote directly for the Labour leader. So it might just mean the British Left was really really far left for a long time and nobody noticed before this.

But the analogy with Trump seems a little too good. People like the same things about both of them. They speak their mind. They don’t care what anyone else thinks. And the establishment obviously hates both.

People have always liked outsiders. But now it’s starting to get ridiculous.

Everyone knows that America is getting more ideologically polarized these days. The right is getting rightier. The left is getting leftier. This puts the Establishment in a bind. The winning strategy had always been to play to the fringe for the primary, then veer towards the center for the general election; the fringe would grumble, but if you played it right you could mollify them and be all things to all people. But now the distance from the fringe to the center is much larger, and proportionally harder to cover without an obvious betrayal. A candidate who wants to get elected on the national level, or even on a local level if the local area is ideologically diverse enough, has to make that betrayal.

Worse, once they’re elected they’ve got to deal with reality. If you try to be too liberal (like raising the minimum wage to $15) or too conservative (like building an immigration wall), then businesspeople with a vested interest in the economy continuing to work start yelling at you, and maybe you back down. There’s this archetypal image of the new President-elect walking into the White House on day one and very serious men in suits telling him “Okay, here are the planks of your campaign platform which don’t have a snowball’s chance in Hell of working and which you are going to drop,” and probably he listens. The clearest example here would be Obama promising to close Guantanamo Bay by 2010, followed by him getting elected and someone asking “Okay, and exactly how are you going solve all of the legal hurdles to doing this?”

The base doesn’t have to worry about reality. If, as people like Robin Hanson suggest, politics is not about carefully selecting policies that most benefit the country so much as about signalling values and ingroup membership, then the base will be interested in enforcing its own particular extremism.

In the past, the center and the fringe were close enough for an uneasy compromise: sure, the base would gripe about “the establishment” and light up at the mention of “an outsider”, but sufficiently canny politicians could still navigate a careful path between their competing demands.

Now that’s becoming harder. The base thinks of the establishment not just as suspicious but as actively hostile; thus the rise of the Tea Party, whose whole purpose was to elect a new kind of conservative who wouldn’t cave in to big government liberals like all the other kinds of conservatives did, and who were a constant thorn in the side of a Republican Party trying to get the most electable people in power so they could do the most electable things.

But now the Tea Party’s actually attained some power, they can’t deny reality any better than their predecessors, and so they keep doing all sorts of crazy things like not shutting down the government over each real or imagined slight. Where do you go from there?

Apparently, you go to Trump.

The most salient feature of Trump – I would say the only salient feature of Trump – is that the establishment hates him. Reince Priebus goes to sleep at night and has nightmares about Trump. The liberal media has important-looking people coming on in suits saying it’s a national embarrassment that anyone could vote for Trump. But in signaling terms, what they’re unintentionally saying is “Moderates hate this guy! He’s too politically incorrect to win over Democrats! Only vote for him if you’re a real Republican.” And Republicans are eating it up. It doesn’t even matter that he’s not that conservative in real life, the media has conducted his campaign for him. Every bad thing the media and the establishment say about him will just make him more popular.

And the same seems true of Jeremy Corbyn.

Warning About Corbyn Only Makes Him Stronger, Warns MP

— Corbyn Warnings (@CorbynWarnings) August 14, 2015

Trace this tendency far enough, and I think it explains why Bernie Sanders is doing better than expected, why Ben Carson has the Republican 2nd place right now, and maybe why Obama won his surprise upset over Hillary in 2008. I predict we are in for a lot more interesting Corbyn-style surprises.

# If You Can’t Make Predictions, You’re Still In A Crisis

A New York Times article by Northeastern University professor Lisa Feldman Barrett claims that Psychology Is Not In Crisis:

Is psychology in the midst of a research crisis?

An initiative called the Reproducibility Project at the University of Virginia recently reran 100 psychology experiments and found that over 60 percent of them failed to replicate — that is, their findings did not hold up the second time around. The results, published last week in Science, have generated alarm (and in some cases, confirmed suspicions) that the field of psychology is in poor shape.

But the failure to replicate is not a cause for alarm; in fact, it is a normal part of how science works.

Suppose you have two well-designed, carefully run studies, A and B, that investigate the same phenomenon. They perform what appear to be identical experiments, and yet they reach opposite conclusions. Study A produces the predicted phenomenon, whereas Study B does not. We have a failure to replicate.

Does this mean that the phenomenon in question is necessarily illusory? Absolutely not. If the studies were well designed and executed, it is more likely that the phenomenon from Study A is true only under certain conditions. The scientist’s job now is to figure out what those conditions are, in order to form new and better hypotheses to test […]

When physicists discovered that subatomic particles didn’t obey Newton’s laws of motion, they didn’t cry out that Newton’s laws had “failed to replicate.” Instead, they realized that Newton’s laws were valid only in certain contexts, rather than being universal, and thus the science of quantum mechanics was born […]

Science is not a body of facts that emerge, like an orderly string of light bulbs, to illuminate a linear path to universal truth. Rather, science (to paraphrase Henry Gee, an editor at Nature) is a method to quantify doubt about a hypothesis, and to find the contexts in which a phenomenon is likely. Failure to replicate is not a bug; it is a feature. It is what leads us along the path — the wonderfully twisty path — of scientific discovery.

Needless to say, I disagree with this rosy assessment.

The first concern is that it ignores publication bias. One out of every twenty studies will be positive by pure chance – more if you’re willing to play fast and loose with your methods. Probably quite a lot of the research we see is that 1/20. Then when it gets replicated in a preregistered trial, it fails. This is not because the two studies were applying the same principle to different domains. It’s because the first study posited something that simply wasn’t true, in any domain. This may be the outright majority of replication failures, and you can’t just sweep this under the rug with paeans to the complexity of science.

The second concern is experimenter effects. Why do experimenters who believe in and support a phenomenon usually find it occurs, and experimenters who doubt the phenomenon usually find that it doesn’t? That’s easy to explain through publication bias and other forms of bias, but if we’re just positing that there are some conditions where it does work and others where it doesn’t, the ability of experimenters to so often end out in the conditions that flatter their preconceptions is a remarkable coincidence.

The third and biggest concern is the phrase “it is more likely”. Read that sentence again: “If the studies were well designed and executed, it is more likely that the phenomenon from Study A is true only under certain conditions [than that it is illusory]”. Really? Why? This is exactly the thing that John Ioannidis has spent so long arguing against! Suppose that I throw a dart at the Big Chart O’ Human Metabolic Pathways and when it hits a chemical I say “This! This is the chemical that is the key to curing cancer!”. Then I do a study to check. There’s a 5% chance my study comes back positive by coincidence, an even higher chance that a biased experimenter can hack it into submission, but a much smaller chance that out of the thousands of chemicals I just so happened to pick the one that really does cause cancer. So if my study comes back positive, but another team’s study comes back negative, it’s not “more likely” that my chemical does cure cancer but only under certain circumstances. Given the base rate – that most hypotheses are false – it’s more likely that I accidentally proved a false hypothesis, a very easy thing to do, and now somebody else is correcting me.

Given that many of the most famous psychology results are either extremely counterintuitive or highly politically motivated, there is no reason at all to choose a prior probability of correctness such that we should try to reconcile our prior belief in them with a study showing they don’t work. It would be like James Randi finding Uri Geller can’t bend spoons, and saying “Well, he bent spoons other times, but not around Randi, let’s try to figure out what feature of Randi’s shows interferes with the magic spoon-bending rays”. I am not saying that we shouldn’t try to reconcile results and failed replications of those results, but we should do so in an informed Bayesian way instead of automatically assuming it’s “more likely” that they deserve reconciliation.

Yet even ignoring the publication bias, and the low base rates, and the statistical malpractice, and the couple of cases of outright falsification, and concentrating on the ones that really are differences in replication conditions, this is still a crisis.

A while ago, Dijksterhuis and van Knippenberg published a famous priming study showing that people who spend a few minutes before an exam thinking about brilliant professors will get better grades; conversely, people who spend a few minutes thinking about moronic soccer hooligans will get worse ones. They did four related experiments, and all strongly confirmed their thesis. A few years later, Shanks et al tried to replicate the effect and couldn’t. They did the same four experiments, and none of them replicated at all. What are we to make of this?

We could blame differences in the two experiments’ conditions. But the second experiment made every attempt to match the conditions of the first experiment as closely as possible. Certainly they didn’t do anything idiotic, like switch from an all-female sample to an all-male sample. So if we want to explain the difference in results, we have to think on the level of tiny things that the replication team wouldn’t have thought about. The color of the wallpaper in the room where the experiments were taking place. The accents of the scientists involved. The barometric pressure on the day the study was conducted.

We could laboriously test the effect of wallpaper color, scientist accent, and barometric pressure on priming effects, but it would be extraordinarily difficult. Remember, we’ve already shown that two well-conducted studies can get diametrically opposite results. Who is to say that if we studied the effect of wallpaper color, the first study wouldn’t find that it made a big difference and the second study find that it made no difference at all? What we’d probably end out with is a big conflicting morass of studies that’s even more confusing than the original smaller conflicting morass.

But as far as I know, nobody is doing this. There is not enough psychology to devote time to teasing out the wallpaper-effect from the barometric-pressure effect on social priming. Especially given that maybe at the end of all of these dozens of teasing-apart studies we would learn nothing. And that quite possibly the original study was simply wrong, full stop.

Since we have not yet done this, and don’t even know if it would work, we can expect even strong and well-accepted results not to apply in even very slightly different conditions. But that makes claims of scientific understanding very weak. When a study shows that Rote Memorization works better than New Math, we hope this means we’ve discovered something about human learning and we can change school curricula to reflect the new finding and help children learn better. But if we fully expect that the next replication attempt will show New Math is better than Rote Memorization, then that plan goes down the toilet and we shouldn’t ask schools to change their curricula at all, let alone claim to have figured out deep truths about the human mind.

Barrett states that psychology is not in crisis, because it’s in a position similar to physics, where gravity applies at the macroscopic level but not the microscopic level. But if you ask a physicist to predict whether an apple will fall up or down, she will say “Down, obviously, because we’re talking about the macroscopic level.” If you ask a psychologist to predict whether priming a student with the thought of a brilliant professor will make them do better on an exam or not, the psychologist will have no idea, because she won’t know what factors cause the prime to work sometimes and fail other times, or even whether it really ever works at all. She will be at the level of a physicist who says “Apples sometimes fall down, but equally often they fall up, and we can’t predict which any given apple will do at any given time, and we don’t know why – but our field is not in crisis, because in theory some reason should exist. Maybe.”

If by physics you mean “the practice of doing physics experiments”, then perhaps that is justified. If by physics you mean “a collection of results that purport to describe physical reality”, then it’s clear you don’t actually have any.

So the Times article is not an argument that psychology is not in crisis. It is, at best, an IOU, saying that we should keep doing psychology because maybe if we work really hard we will reach a point where the crisis is no longer so critical.

On the other hand, there’s one part of this I agree with entirely. I don’t think we can do a full post-mortem on every failed replication. But we ought to do them on some failed replications. Right now, failed replications are deeply mysterious. Is it really things like the wallpaper color or barometric pressure? Or is it more sinister things, like failure to double-blind, or massive fraud? How come this keeps happening to us? I don’t know. If we could solve one or two of these, we might at least know what we’re up against.

# Book Review: Manufacturing Consent

I.

Consider:

It is our view that, among their other functions, the media serve and propagandize on behalf of the powerful societal interests that control…them. The representatives of these interests have important agendas and principles that they want to advance, and they are well-positioned to shape and constrain media policy. This is normally not accomplished by crude intervention, but by the selection of right-thinking personnel and by the editors’ and working journalists’ internalization of priorities and definitions of newsworthiness that conform to the institution’s policy.

[This includes] the ability to complain about the media’s treatment of news (that is, produce “flak”), to provide “experts” to confirm the official slant on the news, and to fix the basic principles and ideologies that are taken for granted by media personnel and the elite, but are often resisted by the general population. In our view, the same underlying power sources that own the media…that serve as primary definers of the news, and that produce flak and proper-thinking experts, also play a key role in fixing basic principles and dominant ideologies.

If I saw this quote on Facebook without attribution, I would assume it was from the latest far-right blog complaining about the liberal media. In fact, it is from Noam Chomsky and Edward Herman’s Manufacturing Consent, which claims that the media acts as lapdog of the dominant neoliberal ideology against leftists of all stripes.

I decided to read Manufacturing Consent because of this basic puzzle: how can both the Left and Right be so certain that the media is biased against them?

Now, in one sense this is not surprising. Everyone believes everything is biased against them. I’ve previously talked about bravery debates, the sort of argument where both sides believe that we’re brave non-conformist speaking truth to power, and they’re toadies of the elite repeating the dominant consensus like sheep. The hostile media effect is a well-known bias where both sides of an issue believe the media is biased against them, even going so far as to both give low fairness ratings to sample documentaries in controlled studies for opposite reasons. So a more general tendency of both sides to accuse the media as a whole of having a hostile agenda is pretty much what we would predict.

The part that surprises me is: I thought that, even objectively, apart from the bias to be expected on both sides, the Right’s case for a hostile media was pretty good. Democrats outnumber Republicans among journalists four to one, and CrowdPAC’s donation analysis rates journalism as among the most liberal professions. There’s an ongoing joke (and some informal analysis) about how disgraced Republicans’ party affiliation is lampshaded and disgraced Democrats’ party affiliation is covered up. And in my own area of interest, it often seems like scientific studies that support liberal beliefs tend to get front-page billing no matter how terrible they are, but scientific studies that cast doubt upon such beliefs them are very rarely mentioned.

And this perception seems to be mirrored by the popular wisdom, where conservatives complain of media bias full stop, and liberals mostly just gripe about Fox in particular.

So Chomsky and Herman’s claim that the media is in fact biased towards conservatives is startling and interesting and deserves a further look.

How exactly do Chomsky and Herman think this media bias works? In Chapter 1, they propose five major mechanisms:

1. The mass media is mostly controlled by large corporations, who therefore support the sorts of things large corporations would be likely to support, like unrestrained capitalism and privileges for the wealthy.

2. The mass media is dependent on advertising, which also involves large corporations who support the sorts of things large corporations are likely to support. Further, these advertisers may have specific interests. For example, Texaco might be less willing to advertise in a source that frequently critiques Big Oil or raises concerns about pollution.

3. Journalists are dependent on sources. The most convenient sources are large well-organized entities in the midst of newsworthy events who issue press releases. For example, by far the easiest source for the latest news about a foreign war is the Pentagon. Furthermore, the Pentagon, while not always in fact trustworthy, enjoys a presumption of trustworthiness; if you interview some random foreigner, you would want to fact-check her very carefully, but if you parrot the Pentagon press release, you are assumed to have done due diligence merely because the source is so official. Other such convenient and official sources of news include the White House, the Department of State, local police forces, and local chambers of commerce. But all of these are members of the establishment and so have a pro-establishment bias. Further, the news relies on “experts” to confirm and comment upon news, and because of incestuous relationships between government, corporations, think tanks, and academia, the most credentialed and salient experts will almost always be pro-establishment.

4. Conservative groups fund “flak machines”, organizations and individuals whose job it is to complain that the media is “biased” whenever they are insufficiently conservative. In these cases, relentless nitpickers will shriek about every slight inaccuracy and condemn the journalists involved as liars and unpatriotic to boot. If the media parrots the official line, then journalists can be almost arbitrarily sloppy and nobody will call them on it. Therefore, journalists who get ground down by the constant harassment will unconsciously shift towards more pro-establishment narratives.

5. Anti-communism is “the dominant religion” of “our cultural milieu” so any journalist who disagrees with the establishment can be smeared with the label “communist” and forced “on the defensive”. Most “have fully internalized the religion anyway, but they are all under great pressure to demonstrate their anti-communist credentials.”

These are interesting ideas, and if supported and developed further they would go a long way towards explaining how the media might have a strong conservative bias despite the liberal leanings of most journalists.

But just after proposing them, the book makes a sudden ninety degree turn to focus on a series of in-depth case studies of US military interference in Third World countries.

The case studies are there for a reason: after Chomsky and Herman establish what they consider to be the true story, they provide examples of the US media consistently misrepresenting even the simplest of facts in ways that flatter the United States government and unfairly malign its foreign enemies. These result in the US getting away with what can only be described as genocide with almost no criticism, even though the facts are plain for anyone to see.

So the idea of media bias hasn’t been exactly dropped. But these studies have disappointingly little relevance to the more general claims that I and presumably most people who bought this book were interested in. Military interference in Third World countries is a very specific subject, and one whose dynamics differ from stories closer to home.

Was I disappointed that the authors didn’t develop their original point about the media more? I was at first. Then I realized this was the book about obscure brutal Third World military conflicts that I’d never known I needed.

II.

Chomsky and Herman are both academics, and they’re both relentless. When they try to prove something, by golly, it stays proved. This is a good thing, in that the book deals with very controversial topics and anything less would be unconvincing. It’s also a bad thing, in that by the ninth or tenth long transcript taken from the same war crimes trial, all of the genocides and village-burnings and nun-rapes start to blend together into a big blob of atrocity, and you can’t remember whether Kouprasith Abhay was the evil generalissimo who launched the pro-US coup and killed thousands, or the good generalissimo who launched the anti-US counter-coup and killed thousands, or the morally ambiguous generalissimo who launched the non-aligned counter-counter-coup and killed thousands.

(his Wikipedia page clarifies that “[his] counter-coup within the counter-coup was ended by the paratroopers responsible for the ongoing coup.”)

But these details are less interesting than the big picture, a sketch of a political system that C&H jokingly term “death squad democracy”.

The general picture is of a third world country that was previously in a fragile social equilibrium. Something disrupts the equilibrium – usually the United States toppling the government because Communists were starting to do well in elections. It is replaced by a weak central government insecure in its power which decides to go after mass movements it perceives as a threat.

The mass movements form guerilla groups to resist government brutality. Supporters of the government form death squads in order to kill suspected guerillas more unethically than the international community would allow the government to do directly. Eventually there is so much violence that anyone who can form a guerilla army and kill their enemies before their enemies kill them does so.

The dictator solemnly declares that what’s going on is a rebellion by communist extremists with associated counter-violence by some grassroots rightist extremists, while he, the dictator, is doing his best to keep the peace. He send in the army, who are secretly or not-so-secretly are also the death squads, and so just make things worse. The United States declares the dictator is a great man who does his best to maintain peace in a troubled nation, and sends him tons of weapons and money. All of these weapons and money mysteriously end up in the hands of the death squads, which of course means the United States has to send in more weapons and money to help the dictator deal with the new threat of these richer, better-armed enemies.

If the dictator is feeling really nice, he will hold an election. The mass movements, communists, and anyone with actual popular support will be banned from participating since they are violent extremists, and the death squads will kill anybody who campaigns against the dictator. The dictator will win the vote handily, and the Free World will declare that since he won the elections, it’s clear that the communists are just violent extremists trying to deny the will of the people and take over for their own nefarious purposes.

This pattern, with slight variation, seems to have happened across the entire Third World at one point or another. Perhaps there will be another coup, and the dictator will be replaced by another dictator, perhaps some foreign country will get directly involved on one side or the other, but the basic logic will not change. For a space of years to decades, tens of thousands of people will be tortured and killed – a few here and there by the communists, but most by the government. Whole villages will be destroyed, freedom of thought will be nonexistent, and everyone except the dictator and a few cronies will be constantly living in fear.

And in a sense, I already knew all of this. We all kind of understand what goes on in banana republics. But for some reason, Manufacturing Consent painted an unusually clear picture that knocked it into relief for me and changed my understanding of a lot of things.

Take, for instance, the second Iraq War. The hawkish position is “we were right to want to remove Saddam, a bad man. We were right to believe that we would win the shooting war quickly and easily. We just couldn’t have predicted the explosion of Sunni-Shiite violence that would erupt afterwards, and that’s not our fault.”

And yet now that I have read Manufacturing Consent, it seems obvious that removing Saddam would cause Iraq to descend into blood-soaked death squads. It is like a law of the universe that Third World countries will descend into blood-soaked death squads at the drop of a pin. Every time the United States has tried to change the government of a Third World nation, the end result has been blood-soaked death squads. Expecting to remove a regime from power without thinking about the blood-soaked death squads seems less like an excusable error and more like missing the very heart of the issue, like expecting to use a nuke without thinking about radiation damage.

But the dove position is almost as bad! It’s “Ha! The hawks thought we would be greeted as liberators! What morons!” This totally misses the point! It’s assuming that if the Iraqis liked us, they would have politely lined up to form a centralized democratic government with a monopoly on the use of force. The problem wasn’t that the Iraqis didn’t like us enough, it was that we did something in a Third World country and expected it not to descend into blood-soaked death squads. That never works.

I am left with a greatly increased respect for the view that it was Western colonialism, broadly defined, that has caused Third World countries all their grief. The problem wasn’t just British people coming in and telling them to work on banana plantations for a while, the problem was the total destruction of the country’s usual rule of law, hierarchies, civic traditions, and social fabric by successive attempts by western-backed dictators to retain power. A couple of decades assassinating anyone who looks out of place and doesn’t do exactly what they’re told, of tearing apart any organization or community that looks strong enough to serve as an alternative to the State or offer resistance – the question is less why Third World countries are so screwed up, and more that they’re not screwed up even worse.

III.

Throughout all of this, the US media could always be counted on to condemn the victims, excuse the aggressors, and totally fail to mention our role in anything.

As per Chomsky, this was rarely done by direct lies, in the form of front page “EVERYTHING FINE IN GUATEMALA, SAY SOURCES”. It was done by a campaign of highlighting certain things, downplaying others, and creating false controversies to cover up the real ones. Their five case studies showcase five different common media biases.

The first study is titled “Worthy And Unworthy Victims”, and compares news coverage of the “worthy victims” killed by America’s enemies to that of the “unworthy victims” killed by America’s allies. The death of worthy victims is treated with outrage, lurid descriptions of every detail of their brutal deaths, intense coverage of every new development in the hunt to bring the killers to justice, focus on the protests their death engenders, and insistence that their death proves a deep and important generalizable lesson about the society in which it occurred. The death of unworthy victims, if covered at all, is treated with “Well, violence sometimes happens, and it’s very sad, but what can we do about it?” Their case study of a “worthy victim” is Jerzy Popieluszko, a Polish priest killed by the Communists; since the Communists were our enemy, we were outraged by the crime. Their examples of “unworthy victims” are the thousands killed in El Salvador and Guatemala, most notably Archbishop Oscar Romero; both countries’ governments were US allies fighting against Communist guerrillas at the time, so their atrocities had to be covered up “for the good of the cause”. As a result, the American populace mostly ended up believing that our enemies were brutal murderers, and our allies were, at best, peace-loving people who were not very good at controlling the violence that always seemed to be breaking out around them.

The second study is “Legitimizing Versus Meaningless Elections”. Most Third World elections are a little sketchy. If the election is in a US ally, it will be covered as a “step towards suffrage in this fledgling democracy”, but if the election is in a US enemy, it will be covered as “a sham” that people are only voting in “for fear of retribution”. The book discusses the elections in Communist Nicaragua versus US-backed El Salvador, showing that by any objective standards the former had fairer, freer elections yet were attacked as a sham by the US media; the latter basically was a sham intended to legitimize a dictatorship, but were praised as a good first step by US media. After reading this chapter it will be very hard for me to take reports of Third World elections seriously again.

The third study is the odd man out, farce in the midst of tragedy. It describes how gullibly the US media accepted the idea of a connection between would-be-Pope-assassin Mehmet Ali Agca and the KGB in the absence of any credible evidence. Yes, C&H admit, Agca did confess to working for the Communists – but only after Italian secret police demanded he do so. Plus he also confessed to lots of other things, including being Jesus Christ, and it was kind of clear that he was a little crazy. In terms of non-psychotic, non-Pope-murdering people who had evidence that the Communists were involved, there was pretty much zilch. But because the Soviets were The Enemy, the media was willing to uncritically pass along anything that discredited them.

The fourth study deals with the Vietnam War, usually considered a case of the media breaking with the establishment and taking a more pacifist, leftist position. C&H argue that this was true only within a very narrow Overton window, where the two acceptable positions were “the US is right to fight for the freedom of South Vietnam” versus “the US is right to fight for the freedom of South Vietnam, but the costs are too high”. C&H argue that nearly everyone in South Vietnam supported Ho Chi Minh except for the dictator and his cronies. The US intervened to save the dictator from his own people, but cast this as saving South Vietnam from North Vietnamese aggression, even though North Vietnam’s involvement was modest. A more honest account of the US role was that they were coming from thousands of miles away to save South Vietnam from “aggression” by the South Vietnamese people. Absent any real enemy except the populace itself, they were backed into a strategy of burning down villages and killing indiscriminately, hoping to keep everyone in such a state of constant terror that they couldn’t do any political organizing. The US media never came close to expressing this position, and therefore at best they could be described as “pro-establishment” and “pro-establishment but sick of losing.”

The fifth study was much like the fourth study, except with Laos and Cambodia. The United States killed about 50,000 people in Laos directly through bombing, and probably more through its consistent support for whichever colonel was launching a coup that day. The US media was completely silent, even though there was ample evidence that it was going on and the foreign media was all over it. Also, when the US media finally got around to talking about it, it was in the context of the supposed “Ho Chi Minh Trail”, whereas most of the bombing was just bombing poor villages in order to deprive the Laotian communists of their natural rural base.

Overall, C&H did a good job of showing ways that the US media could systematically distort foreign wars to cover up the atrocities of US allies, highlight the atrocities of US enemies, and make US actions seem much more noble than the generally chilling evidence would suggest.

IV.

So, do I believe any of it?

C&H are, as mentioned before, really thorough, and they cite everything back and forth twenty ways to Tuesday. But there are ways to be rigorous and dishonest at the same time. C&H had complete control of what incidents to include in their book, and that gives them a lot of power to choose genuinely troubling incidents while not acknowledging any that don’t fit their narrative.

For example, I mentioned before the case of Jerzy Popieluzsko, Polish priest murdered by the Communists. C&H make a big deal on how the US media was saturated with coverage and calls for justice; while they ignored the Salvadorean genocide victims around the same time.

But I notice that the Communists killed about a hundred million people over the course of the twentieth century. Most of these victims did not get the same coverage as Popieluzsko; in fact, we’ve discussed before here how in most cases the media erred on the side of covering these up. Instead of “the media over-covers Communist murders”, it might be “there is wide variance in the media’s coverage of Communist murders, and C&H focused on the most overdone one in order to support their thesis.”

I see this in a lot of places. C&H give a table of various genocides and the news coverage allotted to each. They find that, for example, the news coverage allotted the Kurdish genocide by Iraq (US enemy) was four times greater than the coverage allotted the East Timor genocide by Indonesia (US ally). On the other hand, if they had included Israel in the table, the lesson would have reversed; we hear far more about what Israel (US ally) is doing to the Palestinians than about the Kurds or East Timorese, even though the latter two cases involved far more deaths. Or what if they had included Iran (US enemy)? How many people know about the Iran-PJAK conflict that has claimed almost a thousand lives in the past few years? It’s easy for C&H to cherry-pick examples of well-covered-US-enemies and poorly-covered-US-allies, but it’s not clear that reflects reality very well.

Finally, I’m not sure how much to trust their history. I know very little about the mid-20th century; C&H might be presenting a very one-sided view. The few things I double-checked seem to support this analysis. For example, here’s how they describe Laos in the early 1950s:

A coalition government was established in 1958 after the only elections worthy of the name in the history of Laos. Despite extensive US efforts, they were won handily by the left. Nine of the thirteen candidates of the [communist] Pathet Lao guerrillas won seats in the national assembly, along with four candidates of the left-leaning neutralists (“fellow traveler,” as they were called by Ambassador Parsons). Thus “Communists or fellow travelers” won thirteen of the twenty-one seats contested. The largest vote went to the leader of the Pathet Lao, Prince Souphanouvong, who was elected chairman of the national assembly.

US pressures- including, crucially, the withdrawal of aid – quickly led to the overthrow of the government in a coup by a “pro-Western neutralist” who pledged his allegiance to “the free world” and declared his intention to disband the political party of the Pathet Lao (Neo Lao Hak Sat), scrapping the agreements that had successfully established the coalition. He was overthrown in turn by the CIA favorite, the ultra-right-wing General Phoumi Nosavan. After US clients won the 1960 elections, rigged so crudely that even the most pro-US observers were appalled, civil war broke out, with the USSR and China backing a coalition extending over virtually the entire political spectrum apart from the extreme right, which was backed by the United States.

This seemed so over-the-top cartoonishly evil that I had to check Wikipedia to see if it was an accurate summary. Here’s how they put the same events (editing very liberally for conciseness):

In April, 1953, the Viet Minh’s People’s Army of Vietnam invaded the northeastern part of what was still the French Protectorate of Laos with 40,000 troops commanded by General Vo Nguyen Giap; including 2,000 Pathet Lao soldiers led by Souphanouvong. The objective of the two-pronged invasion was the capture of the royal capital of Luang Prabang and of the Plain of Jars. In November 9 the Pathet Lao began its conflict with the Kingdom of Laos thus beginning the civil war and technically the Second Indochina War while the First Indochina War was still going.

The North Vietnamese invaders succeeded in conquering the border provinces of Phong Saly and Xam Neua, which were adjacent to northern Vietnam and on the northeastern verge of the Plain of Jars. They then moved aside to allow the Pathet Lao force with its mismatched scrounged equipment to occupy the captured ground, and Souphanouvong moved the Pathet Lao headquarters into Xam Neua on 19 April.

On 21 March 1956, Souvanna Phouma began his second term as prime minister. He opened a dialogue with his brother, Souphanouvong. In August, they announced the intention of declaring a ceasefire and reintegrating the Pathet Lao and their occupied territory into the government. However, the Pathet Lao claimed the right to administer the provinces they occupied.

At the same time, they and their North Vietnamese backers ran a massive recruitment campaign, with the aim of forming nine battalions of troops. Many of the new recruits were sent into North Vietnam for schooling and training. This led to United States concern that the Royal Lao Army would be inadequately equipped and trained.

In November, 1957, a coalition government incorporating the Pathet Lao was finally established. Using the slogan, “one vote to the right, one vote to the left to prevent civil war,” pro-communist parties received one-third of the popular vote and won 13 of 21 contested seats in the elections of 4 May 1958. With these additional seats, the left controlled a total of 16 seats in the 59 member National Assembly. Combined with independents, this was enough to deny Souvanna’s center right, neutralist coalition the two-thirds majority it needed to form a government. With parliament deadlocked, the U.S. suspended aid in June to force a devaluation of the overpriced currency, which was leading to the abuse of U.S. aid. The National Assembly responded by confirming a right-wing government led by Phuy Xananikôn in August. This government included four members of the U.S.-backed Committee for the Defence of the National Interest (none of them National Assembly members). Three more unelected CDNI members were added in December, when Phuy received emergency powers to govern without the National Assembly.

Under orders from Souphanouvong, the Pathet Lao battalions refused to be integrated into the Royal Lao Army. Souphanouvong was then arrested and imprisoned, along with his aides. The two Pathet Lao battalions, one after the other, escaped during the night with no shots fired, taking their equipment, families, and domestic animals with them. On 23 May, Souphanouvong and his companions also escaped unscathed.

On 28 July, Communist Vietnamese units attacked all along the North Vietnamese-Lao border. As they took ground from the Royal Lao Army, they moved in Pathet Lao as occupation troops. Poor battle performance by the RLA seemed to verify the need for further training; the RLA outnumbered the attackers, but still gave ground.

On 9 August 1960, Captain Kong Le and his Special Forces-trained Neutralist paratroop battalion were able to seize control of the administrative capital of Vientiane in a virtually bloodless coup, while Prime Minister Tiao Samsanith, government officials, and military leaders met in the royal capital, Luang Prabang. His stated aim for the coup was an end to fighting in Laos, the end of foreign interference in his country, an end to the consequent corruption caused by foreign aid, and better treatment for his soldiers. However, Kong Le’s coup did not end opposition to him, and there was a scramble among unit commanders to choose up sides. If one was not pro-coup, then he had the further decision to make as to whom he would back to counter the coup. The front runner was General Phoumi Nosavan, first cousins with the prime minister of Thailand, Field Marshal Sarit Thanarat. With the Central Intelligence Agency’s support, Sarit set up a covert Thai military advisory group, called Kaw Taw. Kaw Taw, which would support the counter-coup that was mounted; it supplied artillery, artillerymen, and advisers to Phoumi’s forces. It also committed the CIA-sponsored Thai Police Aerial Reinforcement Unit to operations within Laos.

So, things that C&H conveniently forgot to mention: North Vietnam invaded Laos (!), and the Communists gained their power as lackeys for these foreign invaders (!). Although the Communists did well in the 1958 elections, they absolutely did not have a majority in government at the time, and in fact stonewalled the legitimate government. Xananikôn was elected constitutionally by the National Assembly, including the Communists. The Communists refused to stand down their armies and join the national government, and when the government tried to make them, North Vietnam invaded again, with the Communists supporting the foreign invaders. It was in this context that the Neutralists launched their coup, and Phoumi’s CIA-backed countercoup was actually in opposition to it. This is a really different story than C&H’s version. C&H never lie per se, but they leave out things as significant as a giant foreign invasion happening during the middle of the events they’re describing.

Here’s something else I found on Wikipedia: both Chomsky and Herman are considered prominent Cambodian genocide denialists:

Beginning with “Distortions at Fourth Hand”, an article published in the American left-wing periodical The Nation in June 1977, they wrote that while they did not “pretend to know […] the truth” about what was going on in Cambodia during the Khmer Rouge regime of Pol Pot, while reviewing material on the topic then available, “[w]hat filters through to the American public is a seriously distorted version of the evidence available”. Referring to “the extreme unreliability of refugee reports,” they noted: “Refugees are frightened and defenseless, at the mercy of alien forces. They naturally tend to report what they believe their interlocutors wish to hear. While these reports must be considered seriously, care and caution are necessary. Specifically, refugees questioned by Westerners or Thais have a vested interest in reporting atrocities on the part of Cambodian revolutionaries, an obvious fact that no serious reporter will fail to take into account.” They concluded by stating that Khmer Rouge Cambodia might be more closely comparable to “France after liberation, where many thousands of people were massacred within a few months” than to Nazi Germany.

Their book After the Cataclysm (1979), which appeared after the regime had been deposed, has been described by area specialist Sophal Ear as “one of the most supportive books of the Khmer revolution” in which they “perform what amounts to a defense of the Khmer Rouge cloaked in an attack on the media”.[9] In the book, Chomsky and Herman acknowledged that “The record of atrocities in Cambodia is substantial and often gruesome,” but questioned their scale, which may have been inflated “by a factor of 100”. They further asserted that the evacuation of Phnom Penh “may actually have saved many lives,” Khmer Rouge agricultural policies reportedly produced “spectacular” results, and there might have been “a significant degree of peasant support for the Khmer Rouge”: “How can it be that a population so oppressed by a handful of fanatics does not rise up and overthrow them?”

Herman replied to critics in 2001: “Chomsky and I found that the very asking of questions about the numerous fabrications, ideological role, and absence of any beneficial effects for the victims in the anti-Khmer Rouge propaganda campaign of 1975–1979 was unacceptable, and was treated almost without exception as ‘apologetics for Pol Pot’.”

Many other scholars denying or doubting the character of the Khmer Rouge recanted their earlier opinions as the evidence of massive KR crimes against humanity mounted.

They touch on this issue in the book, but I have trouble figuring out what to make of it. Certainly they are outraged that anyone accuses them of denying the Cambodian genocide, and they say this is evil right-wing character assassination propaganda. They then go on to say, kind of flailingly, that also the Cambodian genocide wasn’t that bad, that all the media reports about it were lies, that it was the US’ fault anyway, that the US did worse things anyway, that Cambodia before the genocide was even worse, that America secretly loved Pol Pot and was his best friend, and also shut up shut up shut up. As far as I can get any kind of coherent thesis at all out of this, they seem to be saying they were Gettier cased; every media report of the genocide was a vile right-wing propaganda lie, but coincidentally, a genocide exactly like the one reported in the media occurred.

Herman is additionally criticized for denying the Rwandan and Srebrenica genocides, although Chomsky does not seem to be involved.

And usually I hate terms like “genocide apologist”, because very few people are actually genocide apologists so it’s usually a call to outrage aimed at riling up an angry mob against someone based on one comment they may or may not have said a long time ago.

But in the case where the entire point at issue is a book about genocide scholarship, where the thesis is “everybody else got these genocides wrong, and we are going to tell you the truth about them”, it becomes pretty important if they have a long history of getting genocides wrong.

So I take this book with a grain of salt. I think it treats the topics it covers very rigorously, but (ironically given the subject) the authors’ ability to set the agenda and choose which topics to focus on and which to omit gives them way too much power to shape the readers’ understanding of complex issues.

Do I blame C&H for this? Not exactly. As someone who’s occasionally engaged in some consensus-challenging myself, let me tell you, it’s really hard. Try being perfectly balanced, going out of your way to explain all the facts that disagree with your thesis and pointing out all the grey areas – and no one will listen to you at all. Because if people have heard all their life that A is pure good and B is total evil, and you hand them some dense list of facts suggesting that in some complicated way their picture might be off, they’ll round it off to “A is nearly pure good and B is nearly pure evil, but our wise leaders probably got carried away by their enthusiasm and exaggerated a bit, so it’s good that we have some eggheads to worry about all these technical issues.” The only way to convey a real feeling for how thoroughly they’ve been duped is to present the opposite narrative – the one saying that A is total evil and B is pure good – then let the two narratives collide and see what happens.

And this is really hard, because the same institutions who swallow the utterly bankrupt mainstream narrative whole will suddenly rediscover their skepticism and pick apart every little exaggeration and omission in the contrary narrative. This is the domain of isolated demands for rigor; suddenly no objection is too vague or philosophical, and any amount of emotion or editorializing represents a “bias” that discredits the entire work. So countercultural elements are caught between a rock and a hard place: if they stick to a minimalist stating of the most agreed-upon facts, then it’s not enough to shock people out of their prejudices; any attempt to spin a convincing narrative in the way their mainstream opponents do all the time, and they get attacked for going beyond what can be 100% incontrovertibly defended.

I think C&H handle this impossible balancing act better than most. I think Manufacturing Consent has serious issues with bias, sometimes inexcusably so, but I think its thesis survives these biases. I went into this book with more or less the attitude mentioned above: the classic story of America being great was a bit exaggerated and overenthusiastic, and in fact we did a lot of morally ambiguous things.

I came out of it with more of a primal horror that we spent a lot of the 20th century being moral monsters, and feeling like we have the same sort of indelible black mark on our name as Germany or Russia or Belgium. Whatever factors C&H may have exaggerated, and whatever exculpatory evidence they may have omitted, I doubt that any of it would fully reverse that unpleasant conclusion.

V.

Okay, but what about media bias? Wasn’t that the whole reason we got into this mess?

C&H’s case studies of foreign wars aren’t great tests for their hypothesized mechanisms of bias. Their first two mechanisms are big media corporations pushing a pro-corporate worldview, and big corporate advertisers insisting on programming that reflects well on them and their corporate activities.

And I can see why a mass media dominated by corporate giants might be expected to agitate against labor unions, but it’s harder to see why it is so insistent on covering up a campaign of genocide by pro-American forces in El Salvador. It’s easy to see why they might avoid condemning oil companies in order to preserve ad revenue from Texaco, but harder to see why they would systematically underestimate casualties from US bombing missions on the Plain of Jars in Laos.

Their third mechanism, big Pentagon-style sources with press bureaus, certainly applies very well to these cases. But it doesn’t seem like it should necessarily generalize to every other type of story. When the media is covering an election, or a protest, where is the Pentagon-style source? Although C&H’s point that the police department, etc, can also be sources in this way is well-taken, this seems less pressing for a protest in Seattle than for, say, a bombing campaign in Laos, where a news source might have trouble getting Lao-speaking journalists into the midst of the carnage. Besides, what about cases where this produces the opposite bias? Might newspapers be overly friendly to regulations because they rely upon the regulatory body? What if there is a protest by a large, well-organized group that has cultivated links with the press?

Their fourth mechanism, flak machines, raise a similar issue. C&H view this as a rightist phenomenon almost by definition. They never consider the possibility that, for example, their writing an entire book saying the media is dishonest and biased might count as “flak” on their part. Any conservative criticizing the media is part of a “flak machine” intended to “keep it under control” and “destroy its independence”, but any leftist criticizing the media is bravely trying to expose its biases and bring the truth to light. This seems so obvious to them that they never even have to justify it. This is perhaps understandable in the conflict of foreign wars, where it’s more likely that would-be patriots will condemn reporting that reflects poorly on American troops, but in the context of domestic policy it doesn’t make a lot of sense.

That leaves their fifth mechanism, “anti-communism as the dominant religion of our culture”, a claim which hasn’t aged well since Manufacturing Consent came out in the ’80s. Worse, C&H’s argument for this position is almost word-for-word the same argument that conservatives use to claim that “anti-racism is the dominant religion of our culture”. I’ve even heard them use the specific phrase “dominant religion”.

In their section on “worthy victims” versus “unworthy victims”, C&H describe a certain form of coverage the media reserves for the victims of Communism (section edited for length and clarity):

A. Fullness and reiteration of the details of the murder and the damage inflicted on the victim. The coverage of the Popieluszko murder was notable for the fullness of the details regarding his treatment by the police and the condition of the recovered body. What is more, these details were repeated at every opportunity. The condition of the body was described at its recovery, at the trial when the medical evidence was presented, and during the testimony of the perpetrators of the crime. At the trial, the emotional strain and guilt manifested by the police officers were described time and again, interspersed with the description of how Popieluszko pleaded for his life, and evidence of the brutality of the act…Popieluszko himself was humanized, with descriptions of his physical characteristics and personality that made him into something more than a distant victim.” In sum, the act of violence and its effects on Popieluszko were presented in such a way as to generate the maximum  
emotional impact on readers. The act was vicious and deserved the  
presentation it received. The acts against the unworthy victims [of US anti-Communist client states] were also vicious, but they were treated very differently.

B. Stress on indignation, shock, and demands for justice. In a large proportion of the articles on the Popieluszko murder there are quotations or assertions of outrage, indignation, profound shock, and mourning, and demands that justice be done. Steady and wholly sympathetic attention is given to demonstrators, mourners, weeping people, work stoppages, masses held in honor of the victim, and expressions of outrage, mainly by nonofficial sources. The population “continues to mourn,” “public outrage mounted,” the pope is deeply shaken, and even Jaruzelski condemns the action. The net effect of this day-in-day-out repetition of outrage and indignation was to call very forcible attention to a terrible injustice, to put the Polish government on the defensive, and, probably, to contribute to remedial action.

C. The search for responsibility at the top. In article after article, the U.S. media raised the question: how high up was the act known and approved? By our count, eighteen articles in the New York Times stressed the question of higher responsibility, often with aggressive headlines addressed to that point…

D. Conclusions and follow-up. The New York Times had three editorials on the Popieluszko case. In each it focused on the responsibility of the higher authorities and the fact that “A police state is especially responsible for the actions of its police” (“Murderous Poland:’ Oct. 30, 1984). It freely applied words like “thuggery,” “shameless,” and “crude” to the Polish state. The fact that police officers were quickly identified, tried, and convicted it attributed to the agitation at  
home and abroad that put a limit on villainy. This is a good point, and one that we stress throughout this book: villainy may be constrained by intense publicity. But we also stress the corresponding importance of a refusal to publicize and the leeway this gives murderous clients under the protection of the United States and its media, where the impact of publicity would be far greater.

But of course, that describes to a “t” the media’s coverage of the Ferguson shooting. C&H include a table showing the disproportionate attention given victims of Communism compared to all other types of victims, but the amount of attention given to Ferguson blows all of the Communist murders off the chart.

Does that mean that white policemen fill the same role today that the Soviet Union did back in the 80s? I don’t know. Sure, it’s relevant white policeman killed hundreds of people before Mike Brown with nary a peep from the media. But then, it’s also relevant that Communists killed millions of people before Jerzy Popieluzsko with equally minimal response.

My point is that “anti-Communism” is probably not a uniquely religious belief, and that these “religions” can serve the left as well as the right.

So none of C&H’s five pillars of conservative media domination really seem to stand up very well, which is fine because in their conclusions section C&H switch to a different theory.

They say that the media is a profit-seeking free market, and the best way to get profits is to appeal to advertisers. And the best way to appeal to advertisers is to appeal to the population. And the population wants to hear things that tell them they are good, and their country is good, and don’t challenge or dismay them overly much. Hearing that your government just killed 50,000 Lao civilians is a real downer; hearing that the war on those nasty Commies is going well will keep viewers coming back for more.

But this represents a retreat from the book’s thesis. The media is not exactly a propaganda organ that manipulates the people to serve powerful interests. It’s a tool of the people, giving them what they want to hear – which turns out to be terrible.

And then comes the obvious question – “But, like, fifty percent of the population are liberal, right? Don’t they also get told what they want to hear?”

C&H answer this with the one story that really hammered home the book’s thesis for me: what about Watergate? The media did a great job exposing the lies and corruption of those in power; in fact, of a Republican in power. Does that disprove C&H’s thesis?

No:

The major scandal of Watergate as portrayed in the mainstream press was that the Nixon administration sent a collection of petty criminals to break into the Democratic party headquarters, for reasons that remain obscure. The Democratic party represents powerful domestic interests, solidly based in the business community. Nixon’s actions were therefore a scandal. The Socialist Workers party, a legal political party, represents no powerful interests. Therefore, there was no scandal when it was revealed, just as passions over Watergate reached their zenith, that the FBI had been disrupting its activities by illegal break-ins and other measures for a decade, a violation of democratic principle far more extensive and serious than anything charged during the Watergate hearings.

History has been kind enough to contrive for us a “controlled experiment” to determine just what was at stake during the Watergate period, when the confrontational stance of the media reached its peak. The answer is clear and precise: powerful groups are capable of defending themselves, not surprisingly; and by media standards, it is a scandal when their position and rights are threatened. By contrast, as long as illegalities and violations of democratic substance are confined to marginal groups or distant victims of U.S. military attack, or result in a diffused cost imposed on the general population, media opposition is muted or absent altogether.) This is why Nixon could go so far, lulled into a false sense of security precisely because the watchdog only barked when he began to threaten the privileged.

So for C&H, the media’s rightward bias isn’t “pro-Republican, anti-Democrat”. It’s pro- a conservative establishment in which both Republicans and Democrats collude, and anti- the real left, which it treats as a lunatic fringe too powerless to even be worth mentioning.

This is a new theory, quite different from the five points about corporatism that started the book, and it seems to resolve the paradox of both right and left seeing media bias. The media enforces conformity with the Overton window against both the right and left flanks. Both the rightward and leftward fringes notice the same set of dirty tricks in the media, and describe them in almost exactly the same terms. Thus both sides complain about the other being a “dominant religion”, both sides complain that both major parties are part of the same con, both sides complain that the media restricts debate to a narrow range of acceptable opinion, etc.

And both sides are shouted down in the same terms, too. When the far right complains about the media, academia, and bureaucracy being ranged against them, they get called conspiracy theorists. I myself somewhat hastily made this claim in section 3.2 of my Anti-Reactionary FAQ. More recently, Topher Hallquist makes a similar claim, classily adding that any communities that even dare to associate with people who believe this ought to suffer guilt by association.

Chomsky and Herman are aware of this attack, and begin by saying:

Institutional critiques such as we present in this book are commonly dismissed by establishment commentators as “conspiracy theories”, but this is merely an evasion. We do not use any kind of “conspiracy” hypothesis to explain mass-media performance. In fact, our treatment is much closer to a “free market” analysis, with the results largely an outcome of the workings of market forces. Most biased choices in media arise form the preselection of right-thinking people, internalized preconceptions, and the adaptation of personnel to the constraints of ownership, organization, market, and political power.

And later:

As we have stressed throughout this book, the U.S. media do not function in the manner of the propaganda system of a totalitarian state. Rather, they permit-indeed, encourage-spirited debate, criticism, and dissent, as long as these remain faithfully within the system of presuppositions and principles that constitute an elite consensus, a system so powerful as to be internalized largely without awareness.

I find many smart people, both on the right and the left, say something similar about this same self-organizing consensus enforcement system. Their disagreements about its position seem to be entirely matters of perspective; to a Mexican, America is a northern nation; to a Canadian, it’s a southern one. But despite this substantial agreement and the rivers of ink spilled on the matter, they always describe it in the vaguest of terms, in a style ranging somewhere between “non-technical” and “paranoid”.

If we want to understand politics, I feel like one of the most important subgoals is to figure out the precise ways in which these sorts of alignments arise – in other words, how class warfare solves its coordination problems without most of the people involved being aware of what they’re doing or holding any explicitly sinister thoughts.

I don’t think Manufacturing Consent does much to solve this problem and explain the real nature of the system. But it certainly illuminates one otherwise-easily-neglected corner of it, and offers a window on some of its tricks and on some of the sins it has to answer for.

# Cardiologists and Chinese Robbers

I.

It takes a special sort of person to be a cardiologist. This is not always a good thing.

You may have read about one or another of the “cardiologist caught falsifying test results and performing dangerous unnecessary surgeries to make more money” stories, but you might not have realized just how common it really is. Maryland cardiologist performs over 500 dangerous unnecessary surgeries to make money. Unrelated Maryland cardiologist performs another 25 in a separate incident. California cardiologist does “several hundred” dangerous unnecessary surgeries and gets raided by the FBI. Philadelphia cardiologist, same. North Carolina cardiologist, same. 11 Kentucky cardiologists, same. Actually just a couple of miles from my own hospital, a Michigan cardiologist was found to have done $4 million worth of the same. Etc, etc, etc.

My point is not just about the number of cardiologists who perform dangerous unnecessary surgeries for a quick buck. It’s not even just about the cardiology insurance fraud, cardiology kickback schemes, or cardiology research data falsification conspiracies. That could all just be attributed to some distorted incentives in cardiology as a field. My point is that it takes a special sort of person to be a cardiologist.

Consider the sexual harassment. Head of Yale cardiology department fired for sexual harassment with “rampant bullying”. Stanford cardiologist charged with sexually harassing students. Baltimore cardiologist found guilty of sexual harassment. LA cardiologist fined $200,000 for groping med tech. Three different Pennsylvania cardiologists sexually harassing the same woman. Arizona cardiologist suspended on 19 (!) different counts of sexual abuse. One of the “world’s leading cardiologists” fired for sending pictures of his genitals to a female friend. New York cardiologist in trouble for refusing to pay his $135,000 bill at a strip club. Manhattan cardiologist taking naked pictures of patients, then using them to sexually abuse employees. New York cardiologist secretly installs spycam in office bathroom. Just to shake things up, a Florida cardiologist was falsely accused of sexual harassment as part of feud with another cardiologist.

And yeah, you can argue that if you put high-status men in an office with a lot of subordinates, sexual harassment will be depressingly common just as a result of the environment. But there’s also the Texas cardiologist who pled guilty to child molestation. The California cardiologist who killed a two-year-old kid. The author of one of the world’s top cardiology textbooks arrested on charges Wikipedia describes only as “related to child pornography and cocaine”.

Then it gets weird. Did you about the Australian cardiologist who is fighting against extradition to Uganda, where he is accused of “terrorism, aggravated robbery and murdering seven people”? What about the Long Island cardiologist who hired a hitman to kill a rival cardiologist, and who was also for some reason looking for “enough explosives to blow up a building”?

Like I said, it takes a special sort of person.

II.

Given the recent discussion of media bias here, I wanted to bring up Alyssa Vance’s “Chinese robber fallacy”, which she describes as:

..where you use a generic problem to attack a specific person or group, even though other groups have the problem just as much (or even more so).

For example, if you don’t like Chinese people, you can find some story of a Chinese person robbing someone, and claim that means there’s a big social problem with Chinese people being robbers.

I originally didn’t find this too interesting. It sounds like the same idea as plain old stereotyping, something we think about often and are carefully warned to avoid.

But after re-reading the post, I think the argument is more complex. There are over a billion Chinese people. If even one in a thousand is a robber, you can provide one million examples of Chinese robbers to appease the doubters. Most people think of stereotyping as “Here’s one example I heard of where the out-group does something bad,” and then you correct it with “But we can’t generalize about an entire group just from one example!” It’s less obvious that you may be able to provide literally one million examples of your false stereotype and still have it be a false stereotype. If you spend twelve hours a day on the task and can describe one crime every ten seconds, you can spend four months doing nothing but providing examples of burglarous Chinese – and still have absolutely no point.

If we’re really concerned about media bias, we need to think about Chinese Robber Fallacy as one of the media’s strongest weapons. There are lots of people – 300 million in America alone. No matter what point the media wants to make, there will be hundreds of salient examples. No matter how low-probability their outcome of interest is, they will never have to stop covering it if they don’t want to.

This has briefly gotten some coverage in the form of “the war on police”. As per AEI:

Is there a “war on police” in America today? Most Americans think so, and that’s understandable given all of the media coverage of that topic. A Google news search finds 32,000 results for the phrase “war on cops” and another 12,100 results for “war on police,” with sensational headlines like “America’s War on Cops Intensifies” and “Bratton Warns of Tough Times Ahead Due to ‘War on Cops’.” A recent Rasmussen poll found that 58% of likely US voters answered “Yes” to the question “Is there a war on police in America today?” and only 27% disagreed. But data on police shootings in America that were reported last week by The Guardian tell a much different story of increasing police safety.

According to data available from the “Officer Down Memorial Page” on the annual number of non-accidental, firearm-related police fatalities, 2015 is on track to be the safest year for law enforcement in the US since 1887 (except for a slightly safer year in 2013), more than 125 years ago. And adjusted for the country’s growing population, the years 2013 and 2015 will be the two safest years for police in US history, measured by the annual number of firearm-related police fatalities per 1 million people.

When politically convenient, it is easy to make Americans believe in a war on police simply by better coverage of existing murders of police officers. Given that America is a big country with very many police, even a low base rate will provide many lurid police-officer-murder stories – by my calculation, two murders a week even if officers are killed only at the same rate as everyone else. While covering these is a legitimate decision, it can be deceptive unless it’s framed in terms of things like whether the rate has gone up or down, whether the rate is higher or lower for the group involved than the base rate in the population, and it still seems scary when you explicitly calculate the rate.

But a Chomskian analysis would ask whether the talk of a “war on cops” is really a uniquely bad example of journalistic malpractice, or whether it is bog-standard journalistic malpractice which is unique only in being called out this time instead of allowed to pass.

Let’s stick with coverage of police for consistency’s sake. I’ve made a very similar argument before regarding claims of racist police shootings (see Part D here), but let’s avoid that particular rabbit hole and consider a broader and more unsettling point. We all hear anecdotes about terrible police brutality. Suppose, in fact, that we’ve heard exactly X stories. Given that there are about 100,000 police officers in the US, is X consistent with the problem being systemic and dire, or with the problem being relatively limited?

I mean, it’s hard to say. Quick Fermi calculation: if I can think of about one horrible story of police brutality a week, and assume there are fifty that aren’t covered for every one that is, then per year that makes…

But wait – what if I told you that number was a lie, and there were actually 500,000 police officers in the US? Suddenly the rate of police brutality has decreased five times from what it was a second ago. If you previously believed that there were 100,000 police officers, and that the police brutality rate was shameful but that decreasing the rate to only one-fifth its previous level would count as a victory, well, now you can declare victory.

What if I told you the 500,000 number is also a lie, and it’s actually way more cops than that? Do you have any idea at all how many police there are? Shouldn’t you at least have an order-of-magnitude estimate of what the police brutality rate is before deciding if it’s too high or not? What if I told you the real number was a million cops? Five million cops? Ten million? That’s a hundred times the original estimate of 100,000 – shouldn’t learning that the police brutality rate is only 1% of what you originally estimated (or, going the other direction, 10,000% of that) change your opinion in some way?

(No, I won’t tell you how many cops there actually are. Look it up.)

I feel this way about a lot of things. The media is always giving us stories of how tech nerds are sexist in some way or another. But we may suspect they want to push that line regardless of whether it’s true. How many tech nerds are there? A million? Ten million? How many lurid stories about harassment in Silicon Valley have you heard? Do we know if this is higher or lower than the base rate for similar industries? Whether it’s going up or down? What it would look like if we actually had access to the per person rates?

By now you’ve probably figured out the gimmick, but just to come totally clean – cardiologists are wonderful people who as far as I know are no less ethical than any other profession. I chose to pick on them at random – well, not quite random, one of them yelled at me the other day because apparently contacting the cardiologist on call late at night just because your patient is having a serious heart-related emergency is some kind of huge medical faux pas. I don’t think anyone has ever claimed that there’s any general issue with cardiologists, and as far as I know there’s no evidence for such.

If you read Part I of this post and found yourself nodding along, thinking “Wow, cardiologists are real creeps, there must be serious structural problems in the cardiology profession, something must be done about them,” consider it evidence that a sufficiently motivated individual – especially a journalist! – can make you feel that way about any group.

# Beware Systemic Change

[Epistemic status/edit: After reading comments, no longer sure I agree with Part I. Part II still seems right but possibly a cost that can be outweighed by other factors. I continued to be worried about this without necessarily thinking it is a knock-down argument. I’m still not sure how to balance my support for some systemic change causes against my concern about others. Buck’s comment on morality seems important.]

I.

One of the most common critiques of effective altruism is that it focuses too much on specific monetary interventions rather than fighting for “systemic change”, usually billed as fighting inequitable laws or capitalism in general. For example, Amia Srinivasan, in her review of Doing Good Better,

What’s the expected marginal value of becoming an anti-capitalist revolutionary? To answer that you’d need to put a value and probability measure on achieving an unrecognisably different world – even, perhaps, on our becoming unrecognisably different sorts of people. It’s hard enough to quantify the value of a philanthropic intervention: how would we go about quantifying the consequences of radically reorganising society? […]

Effective altruism, so far at least, has been a conservative movement, calling us back to where we already are: the world as it is, our institutions as they are. MacAskill does not address the deep sources of global misery – international trade and finance, debt, nationalism, imperialism, racial and gender-based subordination, war, environmental degradation, corruption, exploitation of labour – or the forces that ensure its reproduction. Effective altruism doesn’t try to understand how power works, except to better align itself with it. In this sense it leaves everything just as it is.

This same point has been made again and again and again. In response, many effective altruist leaders have gone on to say that they love systemic change and that the movement is entirely in favor of it.

I am not affiliated with the organized effective altruist movement and my opinion has no relation to theirs. They have spent a lot of work trying to convince everyone that they are entirely in favor of pursuing systemic change, I believe them, and nothing I say here reflects on that.

But I, personally, worry a lot about pursuing systemic change.

“Worry about” is not the same as “totally oppose”. This post’s Hansonian title is “Beware Systemic Change” rather than “Against Systemic Change.” But I’m pretty serious about bewaring of it.

First, what do I mean by “systemic change”? Traditional charity, like healing the sick, is almost universally viewed as good or at least neutral. Everyone agrees the sick should be healed; if there are unhealed sick people, it’s because we don’t have the resources to pursue our universally held goal. The same is true of feeding hungry children. It’s true of weird causes like AI risk – some people think it’s silly, but they’re happy to let other people work on them if those people want. It’s even true of things like cutting carbon emissions, sort of. When the Koch brothers say they oppose cutting carbon emissions, they mean they oppose laws mandating such cuts, or budgets that spend communal resources to enforce them. If a private donor offered to pay for scrubbers on every smokestack at zero cost to the rest of the economy, the Koch brothers would have no objection.

Some political issues are kind of like this. People from all over the political spectrum agree that corporate welfare is a bad idea; if we still have corporate welfare, it’s because there’s not enough attention and organization to force politicians to abandon it. In other cases, we all agree something is good but disagree on whether it is an optimal use of resources: for example, most people agree that aid agencies like UNICEF that help children abroad are doing good work, but not everyone agrees with funding them from the federal budget.

Other political issues are not like this. Some people believe that increasing the minimum wage is a laudable goal; other people believe it will hurt the economy or that it violates important moral rights. The reason we don’t have a higher minimum wage isn’t because passing laws costs a lot of money that no one has raised yet, or because no one is paying attention to the issue. It’s because a lot of people oppose it and so far those people are winning, or at least holding their own.

In terms of Freshman English Plot Devices, traditional charity like healing the sick is “man versus nature.” Political issues like the minimum wage are “man versus man”.

When I think of systemic change, I think of man versus man. Even if effective altruists helped governments increase their foreign aid budget, I don’t think Amia Srinivasan and Jacobin Magazine and the rest would think we were participating in “systemic change”. I think at the very least they’d want altruists out in the street demonstrating for higher minimum wages, and at most trying to eliminate global capitalism.

Which is a problem, because a lot of people like global capitalism. A dialogue:

Bob: Man vs. man conflicts raise some thorny issues which man vs. nature conflicts manage to avoid. I think we should be very wary about opening the door to political discussions.

Alice: What? Come on, Bob, you talk about politics all the time!

Bob: I watch porn all the time too; that doesn’t mean I’m proud of it, or think it’s the most good I could do with my resources. Suppose effective altruists get involved in the 2016 US presidential election – which isn’t prima facie a bad idea; think about how easy it would have been to make Gore win in 2000 and how much would have changed if he had. Lots of people work very hard and raise $10 million for the Democrats. Lots of other people work very hard and raise $10 million for the Republicans. Now the Democrats and Republicans are at exactly the same position vis-a-vis each other as they were before the effective altruists got involved, but we have wasted $20 million that could have gone to healing the sick or feeding the hungry. And I’m using money to make things obvious, but the same goes for donating time or advocacy or other resources. If this sort of thing started happening, we would want to promote a general cultural norm of “never spend resources on man vs. man conflicts”. If both sides were equally likely to follow the norm, then the conflicts would remain unaffected but everyone would have more resources to spend on the sick and hungry.

Alice: But effective altruists are very unlikely to donate to both sides of a political issues equally and cancel out. Political views are heavily shaped by demographics, and EAs are likely to skew left just like most other highly-educated groups. Even aside from this, their similar moral assumptions and thinking styles will lead them to converge onto the same side of an issue. Half of Americans are creationist, and almost as many oppose gay marriage, but I would expect fewer than 5% of EAs to hold either position. The idea of donations cancelling out is totally unrealistic. Instead, we should predict that on most issues, most EA donations will go to the same side. We end up not with a wasteful neutrality, but with a large sum of money going to one side, detracted from only slightly by a much smaller sum going to the other. If the cause is important enough, it might still be the most good we could do – the net benefit of (good from giving large amount of money to one side) minus (harm from giving small amount of money to the other side) would still be higher than the benefit of giving all the money to a traditional charity.

Bob: That just kicks the problem one meta-level up. Suppose that on each given issue, effective altruists converge dramatically around one or the other side. If half the time they converge around the right side, and the other half around the wrong side, then over a large number of issues their contributions will gradually even out and sum to zero.

Alice: That’s a ridiculous way of looking at it. We don’t just flip a coin to determine which side to back. We exhaustively study the argument for both sides, the evidence base, et cetera. Then we focus only on those issues where we can be most certain we’re in the right. The odds there are a heck of a lot better than fifty percent!

Bob: But you could make the same argument about picking stocks, couldn’t you? Do lots of research, focus on the ones where you’re most certain that they’re overvalued or undervalued, and then you have great odds of getting rich! But of course, we know that doesn’t work. Everyone else is trying the same thing, and the current position of the stock market reflects the consensus results of that process. You run afoul of the efficient market hypothesis.

Alice: Now you’re just being silly. There’s no efficient market hypothesis for politics!

Bob: But why shouldn’t there be? A lot of people mock rationalists for thinking they can waltz into some complicated field, say “Okay, but we’re going to do it rationally“, and by that fact alone do better than everybody else who’s been working on it for decades. It takes an impressive level of arrogance to answer “Why are your politics better than other people’s politics?” with “Because we want to do good” or even with “Because we use evidence and try to get the right answer”. You’d have to believe that other people aren’t even trying.

Alice: What you’re saying makes a certain kind of sense in Weird Platonic Spherical Cow Perfect Rationality Outside View World. But think about this from the Inside View perspective. Once again, half the country is creationist. Almost half oppose gay marriage. It’s like a stock market where half of the investors are throwing everything they have into the perpetual motion industry. Surely you can admit that even a little bit of intelligence, education, and rationality can actually take you a long way in politics?

Bob: Half the country is creationist, but there’s almost no easy gains from fighting them; any curriculum that federal politics can conceivably affect is evolutionary by this point, and it’s unclear we get any real benefits by going after the last few Alabama middle school students. As for opposing gay marriage, I think you’re going beyond your supposed reliance on evidence here. The strongest conservative case against gay marriage is that it reinforces a centuries-long redefinition of marriage from a strategic partnership focused on child-rearing to a ceremonial acknowledgment of romantic infatuation, potentially leading to a deep shift in the way people think about issues like who to marry, when to have kids, when to get divorced, and how to treat their family. That argument hasn’t been rigorously evaluated by statisticians and found wanting. It’s been found annoying and left untouched. Your differences are foundational assumptions and methodological disagreements about what sorts of issues to focus on, not simple “he made an arithmetic error when calculating the effects” style obvious superiority.

Alice: Really, Bob? You really want to go there?

Bob: Yes. In fact, I worry that this plays into exactly the potential flaws of the effective altruist movement. I can count up all the harms of banning gay marriage: exactly 1.13 million gay people regret not being able to marry, they rate their distress at 3.2/5 on the Likert Scale, that comes out to X QALYs lost per annum, but you have no way of easily quantifying the potential harms of gay marriage, therefore your argument is invalid. A lot of these issues involve trading off easily quantifiable harms on one side versus less quantifiable harms on the other: social trust, cultural cohesion, moral credibility, “freedom” broadly defined, ability to innovate. Highly educated people used to studying science might just be more likely to fall for the streetlight effect and go with the side that promises more quantifiability, rather than the side more likely to be right.

Alice: I…think you’re being deliberately annoying? It seems like exactly the same kind of sophisticated devil’s-advocate style argument we could use for anything. Sure, nothing is real and everything is permissible, now stop playing the Steel Man Philosophy Game and tell me what you really think! It really should be beyond debate that some policies – and some voters- are just stupid. Global warming denialism? Mass incarceration? Banning GMOs? Opposing nuclear power? Not everything is a hard problem!

Bob: I really do sympathize with you here, of course. It’s hard not to. But I also look back at history and am deeply troubled by what I see. In the 1920s, nearly all the educated, intelligent, evidence-based, pro-science, future-oriented people agreed: the USSR was amazing. Shaw, Wells, Webb. They all thought Stalin was great and we needed a global communist revolution so we could be more like him. If you and I had been alive back then, we’d be having this same conversation, but it would end with both of us agreeing to donate everything we had to the Bolsheviks.

Alice: Okay, so the smart people were wrong once. That doesn’t mean…

Bob: And eugenics.

Alice: Actually…

Bob: ಠ\_ಠ

Alice: Fine then. For the sake of argument, the smart people were wrong twice. That still doesn’t…

Bob: It does! A quick run through the history books shows that smart people trying to effect systemic change have an imperfect track record. I won’t say that they’re unusually bad compared to other demographics, but certainly nothing as stellar as the “let’s just not be morons” theory might lead one to expect. You like quantifiable things and specific examples, so let me give you one. I’ve sometimes thought that Friedrich Engels can be considered one of the fathers of effective altruism – at least of the earning-to-give variety. Wikipedia says:

Once Engels made it to Britain, he decided to re-enter the Manchester company in which his father held shares, in order to be able to support Marx financially, so that Marx could work on his masterpiece Das Kapital. Engels didn’t like the work but did it for the good of the cause.

And in one sense, Engels-as-altruist was utterly brilliant. He effectively zeroed in on the most influential thinker of his era, funded an otherwise-impossible level of output from him, and his work directly led to revolutions in a dozen countries with radical change in the lives of billions of people. But in the more important sense, the net effect of his contribution was global mass murder without any lasting positive change. If we count him as an effective altruist – and under the circumstances I’m not sure we can do otherwise – then the net contribution of the movement throughout history has been spectacularly negative. That should make us really concerned. Not “nod sagely and promise to think about it” level of concern, but more “run away screaming” level of concern. That’s why I’m so reluctant to accept your otherwise-reasonable points about the seemingly obvious issues.

Alice: On an emotional level, I get your point. But on a rational level, wouldn’t it be astounding if smart people trying to figure out the safest ways to do the most good consistently made things worse?

Bob: There are many more ways to break systems than to improve them. One Engels more than erases all of the good karma created by hundreds of people modestly plodding along and making incremental improvements to things. Given an awareness of long-tail risks and the difficulty of navigating these waters, I’m not sure our expected value for systemic change activism should even be positive, let alone the most good we can do.

Alice: So don’t go poking around super-complex systems with lots of variables as complicated as “capitalism” versus “communism”. Stick to well-understood things with fairly predictable effects. If we have a little bit more humility than Engels, maybe we won’t fall into the same trap he did.

Bob: All nice and well, except that I do not see even the tiniest sign of supra-Engels levels of humility in the effective altruist movement as it exists today. Recently I have had to deal with lots of our Facebook friends joining and sharing images from a group called “Muh Borders!” which exists to post memes making fun of anyone who opposes Open Borders as a stupid bigot who is not worth talking to:



In terms of “political causes that we can be totally sure won’t backfire and devastate entire countries for generations”, I would place open borders…well, let’s say somewhere in the bottom quartile. A thorough analysis by one of its strongest and most intelligent advocates concludes with “doubt that the American polity could survive and flourish under open borders” but has been mostly ignored in favor of constantly retreading the same old streetlight-illuminated ground of whether immigrants do or don’t affect native wages. And this is the community that is supposed to have solved the hard problem of getting mind-killed by politics, and can now be sure it’s genuinely pursuing the side of Good rather than the side that looks like Good but actually kills tens of millions of people?

Alice: That’s not fair. Yes, there are some people who reflect poorly on the open borders movement, but they’re not all effective altruists, and even humble people who try their best to think about things rationally are allowed to let off steam on Facebook once in a while. The open borders movement has also done a lot of really impressive analysis, and even though there are risks, given the potential benefits it really can be thought of as a no-brainer.

Bob: I am a coward and will stick to buying bed nets, thank you.

II.

There’s another problem with man vs. man: the people we want to recruit are men, and the people we want to make our movement out of are men.

(That came out sounding more sexist than I intended. You know what I mean.)

Several people have recently argued that the effective altruist movement should distance itself from AI risk and other far-future causes lest it make them seem weird and turn off potential recruits. Even proponents of AI risk charities like myself agree that we should be kind of quiet about it in public-facing spaces.

As someone whose own views on open borders are mixed (I should probably write a post), I am really turned off by memes like the one above. And since only seven percent of Americans fully support open borders, that’s a lot of potentially turned-off people. They’re going to go on effective altruist sites, see that a big part of the movement is arguing for a policy that they abhor, and notice their potential colleagues talking about how people like them who oppose that policy are stupid and parochial and hate foreigners. “We think you’re wrong and stupid, come join our movement” makes a really crappy recruiting pitch. But it is the pitch we are sending to anyone who isn’t a Silicon Valley libertarian, George Mason University economics professor, or Vox.com journalist – the only three groups from which I have seen a level of open borders support much beyond the lizardman level.

If effective altruists are split on an issue, then they’re working at cross-purposes, like the people above who donated $20 million to both the Democratic and Republican campaigns. But if effective altruists are not split on an issue, then they’re projecting a unified Effective Altruist Consensus on it which is going to look pretty intimidating to anybody who disagrees. And if there are enough of these issues, then a randomly selected person is almost certain to disagree with at least one of them. The more different from the EA stereotypical demographic they are, the more likely such a disagreement will be. Politics is the mind-killer and quickly takes over from everything else; I do not think political disagreements can stay quiet and harmless for long. If two people are both committed to healing the sick and feeding the hungry but one believes in open borders and the other in a more Bernie Sanders style approach to immigration – not even conservative, just a Bernie Sanders style approach! – they can peacefully coexist in an effective altruist movement focusing on traditional charity, but one focusing on systemic change is likely to get pretty heated.

If you think this is overly pessimistic, think back to the issues with the most recent EA Summit, which advertised fully vegetarian meals but added non-vegetarian options at the last second. This became a big enough scandal that I, who was two thousand miles away from the conference, got inundated with arguments about it on Facebook, Tumblr, and this blog. Several people threatened to quit effective altruism entirely, though I don’t know if any of them followed through.

This is a community that can literally almost tear itself apart over the question of what to have for lunch. I think there might be too much dynamite around to risk shooting off sparks.

And I also think effective altruism has an important moral message. I think that moral message cuts through a lot of issues with signaling and tribal affiliation, that all of these human foibles rise up and ask “But can’t I just spend my money on – ” and effective altruism shouts “NO! BED NETS!” and thus a lot of terrible failure modes get avoided. I think this moral lesson is really important – if everyone gave 10% of their income to effective charity, it would be more than enough to end world poverty, cure several major diseases, and start a cultural and scientific renaissance. If everyone became very interested in systemic change, we would probably have a civil war. Systemic change is sexy and risks taking over effective altruism, but this would eliminate a unique and precious movement, in favor of doing the same thing as everybody else.

If effective altruism became more political, it would likely fade seamlessly into something like the Brookings Institution (a top-tier think tank whose $100 million yearly budget is by my calculations well above what the entire world combined spends on deworming) or the Cato Institute (another top think tank whose $30 million budget is likely more than all AI risk charities and all effective animal rights charities combined). Probably the staff of the Brookings Institution go into work each day thinking “How can I best improve the world by giving it better policies?”, and I admire that, but they don’t have the same sort of moral mission as effective altruism and it would be disappointing to see the latter collapse into an annex of the former.

# Vegetarianism for Meat-Eaters

[Content warning: discussion of animal suffering. If you don’t care about animal suffering, this post is probably not for you. There is no reason to read it anyway and loudly complain in the comments.]

Brian Kateman on QZ.com writes that We Need More Meat-Eating Animal Rights Activists. Finally, the mainstream media gives me ex cathedra permisson to say things that are kind of hypocritical!

I believe animals probably have moral value. I also eat meat. There is obvious tension between these positions; animals suffer and (obviously) die during meat production. I can only say in my defense that I tried being a vegetarian for several years and it was horrible and I ended up subsisting almost entirely on bread and Quorn and I don’t want to go back there.

But over the past few years I’ve read about two ideas that have changed the way I look at meat-eating and significantly reduced my moral footprint with minimal inconvenience. These are not original to me and I don’t take credit for them, but I hope that the people involved won’t mind me taking this advantage to publicize them more widely.

1. Eat Beef, Not Chicken

This argument is so simple I feel dumb for not thinking of it myself; instead, I take it from Julia Galef and Brian Tomasik. Suppose I get about a third of my daily calorie requirement from meat; that adds up to 250,000 calories of meat a year. Further suppose that it’s split evenly between 125,000 calories of beef and 125,000 calories of chicken.

The average cow is very big and makes 405,000 calories of beef; the average chicken is very small and makes 3000 calories worth of chicken. So each year, I kill about 0.3 cows and about 42 chickens, for a total of 42.3 animals killed. [1] [2]

Suppose that I stop eating chicken and switch entirely to beef. Now I am killing about 0.6 cows and 0 chickens, for a total of 0.6 animals killed. By this step alone, I have decreased the number of animals I am killing from 42.3/year to 0.6/year, a 98% improvement.

The difference becomes even bigger once you compare levels of suffering. Chickens are probably the most miserable farm animals; they are mutilated, packed into tiny cages to the point of immobility, left to fester in their own waste, and bred so intensively for size that their bodies cannot support them and they likely experience severe musculoskeletal pain. Although cows’ lives are also pretty terrible too, Brian Tomasik estimates that chickens’ suffering is about twice as bad. Taking this into account, switching from 50-50 to all-beef reduces your contribution to animal suffering as much as 99%. [3] [4] [5]

I find that I’m indifferent between beef and chicken as far as taste, so this is a no-brainer for me. The few times I’m making a recipe that really, truly, can only be done with something sort of chicken-like, Beyond Meat vegetarian fake chicken strips are an almost-tolerable substitute.

2. Use Ethics Offsets By Donating to Animal Charities

[EDIT 12/30/16: SERIOUS PROBLEMS WITH THIS SECTION, SEE MISTAKES PAGE FOR DETAILS]

I talked about this before in Ethics Offsets, but I think the original argument comes from Katja Grace.

Animal-related charities are very effective. Animal Charity Evaluators, a sort of animal version of GiveWell, lists really really impressive impacts for small donations:

Animal Equality: 11 animal lives saved per dollar  
Mercy For Animals: 9 animal lives saved per dollar  
Humane League: 3 animal lives saved per dollar

These numbers are high, but not impossibly so. For example, the Humane League spent about $50,000 convincing school districts to switch to cage-free eggs and have “Meatless Mondays” at their cafeterias; this resulted in about 3.2 million fewer meat-containing lunches, meaning several hundreds of thousands of chickens saved.

Okay. If you followed the advice in Part 1 and switched to beef, you’re currently killing 0.6 animals per year. If you donate six cents per year to animal-related charities, you’re animal-neutral. Donating $0.06 sounds…a lot easier than being vegetarian for a year? [6]

Or donate $60, and save more animals than an entire village full of vegetarians. At this point it’s starting to look like maybe personal vegetarianism is more of a symbolic/non-consequentialist decision in comparison, and a meat-eater with a little pocket change to spare can bask in near-unlimited moral superiority even to their most scrupulously vegan friends. Is this too good to be true?

One reason it might be too good to be true is that Animal Charity Evaluators is overly optimistic. But it would be really hard for their optimism to change this strategy substantially. Suppose that they were off by an order of magnitude, and you only save one animal per dollar. You can still offset an entire year’s beef-eating for $0.60. Even if they’re off by three orders of magnitude and it takes $60 to offset a year of eating beef, most people would probably still rather pay sixty bucks than become vegetarian.

A more serious complaint is that this strategy is hypocritical or self-defeating. After all, it looks like most of the gain from these charities comes from convincing other people to be vegetarians. From a Kantian point of view, “try to get other people to become vegetarian without being one yourself” isn’t universalizable; if everyone did it, there would be nobody to actually be the vegetarians! Is it ethical for non-vegetarians to try to spread vegetarianism among other people? Here are four arguments that it is:

First, consequentialism. From a consequentialist point of view, “is it okay to cause a good thing to happen even if…” always gets answered yes. Do you save the animals? Yes? Then what’s the problem? The true consequentialist doesn’t even understand the question.

Second, these charities don’t necessarily demand people become full vegetarians. They may recommend that people cut down on the amount of meat they eat, or switch from chicken to beef as in Part 1, or support laws enforcing more humane living conditions for farm animals. Some evidence supports asking meat-eaters to cut down on meat as the most effective form of animal outreach. A non-vegetarian who has taken some of these steps themselves can support these without worrying about hypocrisy.

Third, your situation is not necessarily the same as other people’s situations. One reason I’m not a vegetarian is that I really really hate vegetables. Other people might love vegetables and just need a little push to have more of them. I can endorse that people become vegetarian if it is easy for them without necessarily endorsing vegetarianism for myself.

Fourth, and I think most important, the economics check out. Instead of universalizing the principle “become vegetarian”, suppose we tried to universalize the principle “find some way to be animal-neutral,” that is, live your life in such a way that on net you are not killing animals. And suppose everyone knew there were two strategies for doing this: either become vegetarian yourself, or offset your lifestyle by donating to advocacy organizations that convert other people to do so.

And suppose that, upon hearing that it only takes a $60 donation to offset their lifestyles, 90% of people choose the donation rather than the personal conversion. This makes the cost of outreach go up. That is, when I donate my $60, the advocacy organization uses it to convert Alice, who decides to donate $60 herself, which the advocacy organization uses to convert Bob, who decides to donate $60 himself, which the organization uses to convert Carol…and so on to the tenth person, who finally decides to become vegetarian themselves. If this happened, our premise that it takes the charity $60 to convert one new vegetarian would be false. In fact it takes them 10 donations of $60, or $600.

As long as people know that they have the option of offsetting via donation, the possibility that people would rather donate than become vegetarian themselves is priced into the cost of the offset. That means that if the cost of an offset is currently $60, it’s because we’re hitting people for whom $60 is genuinely their reserve price; they prefer becoming vegetarian to paying a $60 offset (probably for moral/symbolic reasons). These people are low-hanging fruit; once they’re exhausted, the offset price will rise, and people for whom vegetarianism is only a mild inconvenience will find themselves preferring to become vegetarian themselves rather than paying. Once even the middle-hanging fruit is exhausted, the price of the offset will be prohibitive and only the people for whom vegetarianism is an extraordinary inconvenience will continue to take that route. Once there are no more potential vegetarians left to convert, the offset cost will become the cost of saving animals via political action, improved technology (eg cultured meat), or changes to farming conditions.

This dynamic becomes even more interesting if you add the (unjustifiable but interesting) assumption that anyone not becoming vegetarian themselves is required to offset their choice by converting two other people to vegetarianism. Then you get a sort of virtuous Ponzi scheme which ends with a lot of vegetarians (albeit not necessarily in a reasonable amount of time).

I try to donate some money to an effective animal charity each year, above and beyond what I’ve pledged to donate for other reasons, in order to compensate for the remaining meat I refuse to cut out of my diet.

Footnotes

1. I use the term “kill” because it’s a simple way of looking at things, but most of the moral cost of eating meat is causing the animals to spend years living in terrible suffering on factory farms. The actual killing is probably a mercy in comparison. When I say that something “prevents forty animals from being killed”, the longer and more accurate version might be “prevents forty animals from coming into existence, suffering intensely, and then being killed”. This does raise some more philosophical questions like whether it’s better to live a life of terrible suffering than to never be born at all, but I’m really comfortable answering that one with “no”.

2. This same argument comes out against eating other small animals like fish. Although in theory wild-caught fish ought to live okay lives and potentially be more ethically acceptable than farm-raised animals, given limited wild-catching ability each wild-caught fish eaten may deplete a fixed number of them and push other people to eat farm-raised fish instead.

3. Eggs raise some of the same issues as chickens, and Julia Galef suggests eggs are one of the worst things you can eat. I think her assessment is pessimistic; eggs are terrible on a calorie-for-calorie basis, but if we’re talking about which animal products to urge people to give up, this is counterbalanced by nobody except Gaston getting too many calories from eggs. Someone who eats one egg with breakfast every day kills about one chicken a year; somebody who has a chicken dinner every other night kills about forty chickens a year. Although egg chickens probably lead worse lives than meat chickens, the difference isn’t overwhelming. Avoiding incidental egg consumption like the eggs in baked goods is hard and probably not the highest-value pro-animal intervention given the low number of eggs involved.

4. This analysis neglects consideration of whether cows, being bigger-brained and more “evolutionarily advanced” than chickens or fish, might have greater moral value. I don’t know how to deal with that question, except that it would surprise me if they had more than forty times the moral value.

5. The existence of supposed humane animal products (“Free range eggs!” “Pasture-fed cows!”) complicates this a little bit. The unanimous opinion of people who know about this sort of thing is that free range eggs are kind of a scam; regulations only specify that these chickens must have “access” to the outdoors, but farmers exploit the letter of the law to cram thousands of chickens into industrial barns with a single tiny door to a couple-square-foot cement yard that the overwhelming majority of the chickens never even see. “Cage-free” chickens or eggs seem probably better than the alternative but still pretty horrible. “Pasture-fed beef” usually does involve a pasture in some way and is not a total scam but is probably not as nice as you would think. I try to buy pasture-raised free-range cows, and I think that the slightly higher standards of humane beef over humane chicken make another good argument in favor of beef consumption, but I try not to fool myself into thinking that this decision alone goes especially far.

6. If you also eat chicken, the offset cost rises to $4.

# The Problems With Generic Medications Go Deeper Than One Company

I.

Like many people, I recently read about Turing Pharmaceuticals’ purchase of anti-toxoplasma drug Daraprim and subsequent price increase of 5000%. Vox and Marginal Revolution have already done some good work addressing this particular case, but have only touched upon the broader issue: that everything about generic medications is approximately this terrible.

As far as I’m concerned, the interesting aspect of this case isn’t just that the CEO of Turing is an asshole who is lining his own pockets with zillions of dollars by gouging AIDS patients. I assume most pharmaceutical company CEOs are assholes who would line their own pockets with zillions of dollars by gouging AIDS patients if the opportunity presented itself. The interesting aspect of this case is that the CEO of Turing got the opportunity. How?

In the United States, pharmaceutical companies that discover a new drug are granted a 20-year term of exclusivity to reward them for the public service of drug research. During this time, they can and do price-gouge as much as they want. After twenty years, the drug becomes public domain and anybody who wants can compete to produce it, usually leading to a precipitous fall in costs. But Daraprim is fifty years old; its patent is long-since expired. So Sarah Kliff from Vox asks the obvious question: why doesn’t someone just produce a competitor?:

Daraprim isn’t a frequently used drug. The New York Times estimates that between 8,000 and 12,000 prescriptions get filled annually. You could only fill about a quarter of a baseball stadium with the number of people who take the drug in a given year.

So think about a generic drug manufacturer looking at the Daraprim situation. There are fixed costs associated with building a new plant (or possible lost revenue on other drugs, if they switch production at an existing plant), getting samples of the drug, and figuring out how to make the generic product…with Daraprim, there simply isn’t a big enough patient population for a competitor to sell a “good amount” to. And this is, more generally, a problem with the markets for drugs that only a small number of patients use. They often aren’t big enough to support two competitors.

Moreover, there’s risk associated with starting a drug price war. Let’s say I decide to launch Sarah’s Generic Drug Company, and I’m pretty sure I can break even by slightly undercutting Turing and charging $700. What happens if Turing responds by dropping its price down to $500, or even back to $13.50? It will keep all its patients — and my nascent drug company is likely going bankrupt.

This is definitely part of the story. On the other hand, what about Longecity group buys? Someone on a drugs forum hears about a cool experimental chemical that sounds fun to try. They get a couple dozen friends in on it and pay a lab in China a few hundred dollars to synthesize a big batch. Then the Chinese ship it over, they distribute it to their friends, and they all get a decent supply of a totally novel drug for a few dollars a pill – compared to the $750 per pill that Turing is charging for Daraprim. I am not a chemist, but the Daraprim molecule does not look very intimidating. I bet if a group from Longecity got a couple of toxoplasma patients together for a group buy, they could all get treatments for maybe a few hundred dollars each instead of the $63,000 Turing is now charging. In fact, I encourage somebody to do exactly that as an act of civil disobedience/political activism and win themselves some free publicity.

So how come Longecity can do this, but real generic pharmaceutical manufacturers can’t? I’m not totally sure, but my best guess is that it involves bioequivalence studies (different from purity studies). Generic drugs don’t need the excruciatingly drawn-out safety and efficacy studies required of new brand-name medications, but they do need to pass a bioequivalency study proving that their drug is absorbed the same way as the original. According to Wikipedia, the most common type of bioequivalence study is to “measure the time it takes the generic drug to reach the bloodstream in 24 to 36 healthy volunteers; this gives them the rate of absorption, or bioavailability, of the generic drug, which they can then compare to that of the innovator drug”.

This might not seem so bad, but it must be harder than it sounds. This site, whose style is overly bombastic but whose information seems mostly correct, says that:

The cost and time involved in the ANDA [generic application] process varies depending on the drug, its safety, how long it has been on the market, etc. To have an ANDA approved, it typically requires an investment of about $2 million, and it takes a total of two to three years to get the drug to market…in addition to these costs, a company should budget 15% for legal fees, because wherever there is a big manufacturer with a sizable market share involved, they will sue, just to try to eliminate more competition from the market.

This adds an important extra dimension to Vox’s theory that it’s just too hard to start making a generic medication. If all you want to do is synthesize an active ingredient in powder form, and you’re not too concerned about staying on the right side of the law, it costs pennies and takes however long you need to FedEx something from China. If you also want FDA approval, it costs $2 million and takes two years.

Remember, Daraprim is used by about 10,000 people per year, and before the recent Turing price markup, it cost $13.50 per pill x eighty pills per treatment. 10,000 \* 80 \* $13.50 = about $10 million per year, of which maybe $5 million was profit. That means you have to capture a big chunk of the Daraprim market before it’s worth trying to get yourself approved to make Daraprim; the FDA is essentially telling pharma companies to “go big or go home”. Nobody wanted to go big, so they all went home.

In the absence of this barrier, it would be easy for small boutique companies with a couple of chemical engineers on hand to spend a few weeks manufacturing a few thousand doses of the drug whenever it was necessary to meet demand. This is how the supplement and nootropic industries work right now, and nootropics are dirt cheap, even though a lot of “nootropics” are the same chemicals as regular expensive medications except with a “not intended for human consumption” label slapped on the bottle that everyone knows to ignore.

I think this might be what’s going on with generic modafinil. Last week I prescribed some modafinil to one of my patients and got a call back from their insurance company saying it was denied because it cost too much.

I told the insurance company that was silly because modafinil only cost about $60 a month.

The insurance company said no, it cost way more than that.

This surprised me, because half the rationalist community uses modafinil, and even some of the doctors I work with use modafinil on long night shifts, and they all get it for $60 a month from places like ModafinilCat.

But according to Nootriment, a month’s supply of modafinil at real bricks-and-mortar pharmacies costs anywhere from $469.23 (Costco) to $850.84 (RiteAid). I’m not totally sure what’s going on, but my guess is that ModafinilCat (illegally) buys it from people who haven’t gone through the FDA’s bioequivalence testing, and RiteAid buys it from people who have. As far as I can tell, both are made by Indian pharmaceutical companies unrelated to the original American company who discovered the drug, but RiteAid’s Indian pharmaceutical company has put more work into staying on the right side of the US government.

If any of my patients are reading this and are upset because I prescribed them a drug which they couldn’t afford, I unreservedly apologize. I was laboring under the misapprehension that the pharmaceutical market made sense.

II.

No tour of terrible generic medications policies would be complete without a stop for Kesselheim and Solomon’s analyis of the Unapproved Drugs Initiative of 2006.

The FDA wanted to encourage people to study drugs that were already in the public domain and get them up to FDA standards. This is potentially a very noble plan. I’ve written before on how it’s basically impossible to get melatonin to interface with the health care system because it got into the public domain without the relevant FDA standards being met. Likewise, there’s no interest in using minocycline to treat schizophrenia because it’s a public-domain drug and nobody profits off of doing the FDA compliance work. So the FDA was definitely responding to a real problem.

Their solution, though, was to say that if anybody did a good enough study on a public domain drug, they could grab it out of the public domain and have it be their exclusive drug for the next while. This was a terrible terrible terrible idea.

Colchicine is a very popular and very effective gout treatment extracted from the Colchicum plant. It’s been used for so long that its first recorded mention in medical literature is on an ancient Egyptian papyrus. The medievals called it “hermodactyl”; Arabic physician Avicenna recommended it; notable gout sufferer Ben Franklin brought the first Colchicum specimens to North America.

But the ancient Egyptians, being a primitive and barbaric people, had no FDA. And although many different groups had done studies proving colchicine effective, none of them had done so on the official FDA forms. In 2007 a company called URL Pharmaceuticals did an official FDA safety study, showed that yup, it was safe all right, and for this service were granted exclusive right to produce colchicine. After suing all other colchicine producers out of business and establishing a monopoly, they raised the price of colchicine by 5000%, costing gout patients thousands of dollars a year.

According to FiercePharma, something similar happened with hydroxyprogesterone caproate, although the FDA later changed its mind. I can’t find any other examples, but the legal framework is still there if someone else wants to try.

III.

Other times generic manufacturing proceeds smoothly. A drug is popular and many different pharmaceutical companies pass the bioequivalency tests, get in on the action, and compete with one another. Nobody snatches it out of the public domain at the last second and receives a new monopoly on it. The companies are able to sell it to the pharmacies for a reasonable cost.

Now you get to have a completely different set of things go wrong.

Michigan Drug Prices is my state’s official drug price register. You can type in any Michigan ZIP code and any drug and find out how much it costs at all your local pharmacies. It’s pretty neat.

Celexa has been generic for more than a decade, it’s got a reputation for being inexpensive, and I prescribe it a lot. Let’s see how much my patients have to pay.

The closest RiteAid to my office charges $4 for a 30-day supply of Celexa 20 mg. The local CVS sells the same amount for $19.79. The local Walgreens sells it for $24.99. And the local KMart will sell it to my patients for the low, low cost of $88.15. That’s an…interesting…range of prices.

If I try to buy it off GoodRx.com, a site that offers pharmacy price comparisons, I can get it for $3.60 from a mail-order pharmacy. But I can also get it for $6.64 from K-Mart, special offer for GoodRx customers only. $10.00 from Walgreens. $11.99 from CVS. All the same stores that were trying to gouge me before. As soon as you take the basic step of saying “by the way, I’m also comparing costs with other pharmacies” their prices drop 90%.

I am far from the only person to notice this. PBS did a segment on one of the reporter’s mothers looking for a breast cancer drug. She originally paid $400 a month for it, which is steep but perhaps worth the cost as a high-tech treatment for a potentially fatal illness. Then she went to Costco and found the same medication cost $10.

Why does this sort of thing happen? I’m not sure. I expect it has something to do with insurance co-pays; if an insurance looks at some kind of average cost of Celexa and decides that the Celexa co-pay will be $5, then it doesn’t much matter to the customer whether they buy it from a pharmacy charging $10 or $10,000. But why doesn’t the insurance company do one the thing everyone in health care agrees insurance companies do best: send whiny faxes complaining that they’re not going to pay you? I don’t know.

But for now you might want to try using something like GoodRx.com if you’re buying expensive medications. And stay away from cats, because there’s never been a worse time to get toxoplasma.

# Swifties 3: The Race Is Not To The Swifty

[see: Wikipedia: Tom Swifties, Tom Swifties Written By An Author Willing To Go To Any Lengths To Make A Tom Swifty Thus Resulting In Constructions That Often Require More Work For Readers Than For The Author, Fifty Swifties, and Fifty More Swifties. Previously on Twitter here. Some of these are from the comments to the last post.]

1. “She eventually absorbed so much radiation that her bottom half mutated into a fish’s tail,” Tom said mercurially.

2. “Stay away from nuns,” Tom said conventionally.

3. “Back during Late Antiquity, everyone lived in fear of Attila and his hordes,” Tom said a hundred times.

4. “It said he was eaten by a bare, so either that’s a typo or he was devoured by the act of exposing something,” Tom said verbatim.

5. “You’ll have to stand,” Tom said deceitfully.

6. “Little plays are such a useful way to teach children good behavior,” Tom said schizoaffectively.

7. “…” Tom said immutably.

8. “I’m an only child,” Tom said in unison.

9. “Look, a Confederate general!” Tom said icily.

10. “Why yes ma’am, I AM the Tom from those Twitter one-liners you’ve heard,” Tom said pungently.

11. “I’m not going to make a deathbed conversion,” Tom said diagnostically.

12. “I’m using behavioral conditioning to train lions to keep quiet,” Tom said to Rorschach.

13. “I used to be a priest, but I was defrocked for an improper relationship on the job,” Tom said inundated at work.

14. “I’m here helping people displaced by the earthquake,” Tom said with intensity.

15. “We’ve been pinned underneath fallen logs,” Tom said treasonously.

16. “I went rock-climbing with my girlfriend,” Tom updated.

17. “The defibrillator worked!” Tom said, repulsed.

18. “My karate instructor died,” Tom said, desensitized.

19. “Godzilla, I can’t believe you devoured part of South Africa,” Tom transvaluated.

20. “I was running late today, so I had lunch in my cubicle,” Tom incubated.

21. “But they dug too greedily, and too deep,” Tom undermined.

22. “The new environmental regulations will make mineral extraction less profitable,” Tom said, determined.

23. “He’s sleeping six feet under now,” Tom said depressed.

24. “I guess I lost the genetic lottery,” Tom said, drawing a portrait.

25. “SKULLS FOR THE SKULL THRONE!” Tom said, skulkingly.

26. “I’ve gotten 0.028 countries to join together in a political and economic union,” Tom said in his milieu.

27. “For here I am, sitting in a tin can, re-entering Earth’s atmosphere,” Major Tom said incandescently.

28. “The mailman just left my mail on the dirty ground?! Really?!” Tom said postindustrially.

29. “I’m writing a book based on ‘The Tell-Tale Heart’, except instead of a horror story it’s a comedy,” Tom said politely.

30. “Is the guy in that coffin Dracula, or just an ordinary corpse?” Tom countermanded.

31. “I think China has enough foreign exchange reserves,” Tom said for example.

32. “Every time the server goes down, I have a Norse god zap it with lightning to get it back up,” Tom said with authority.

33. “Help, I’ve been buried alive!” Tom engraved.

34. “I’ll never be an A-list celebrity” Tom berated. (source)

35. “If you were any good you’d have the Ambassador’s job,” Tom said disconsolately. (source)

36. “Germany should exit the Eurozone” Tom remarked. (source)

37. “Maybe he was knighted for his contributions to Austrian economics,” Tom surmises.

38. “We should give the Western US back to the Native Americans,” Tom said unsettlingly.

39. “I’m not going to give that jerk Procrustes the satisfaction,” Tom said self-defeatingly. (source)

40. “This new-ideas conference has sure gotten effeminately quaint.” Tom tweeted. (source)

41. “Everyone’s date of birth is in 2007,” Tom said alternatively. (source)

42. “Weasley for president!” Tom said electronically. (source)

43. “Let the other guy take the paddle,” Tom said heroically. (source)

44. “Let’s make a deal – I’ll stop doing sit-ups if you do,” Tom said abstrusely. (source)

45. “My former wife mentioned me in her newest paper,” Tom said excitedly. (source)

46. “How Can Mirrors Be Real If Our Eyes Aren’t Real?”, Tom asked unreliably. (source)

47. “Your hair looks terrible,” Tom said distressingly. (source)

48. “I’ve stolen the treasures of the Shrine of the Bab,” Tom said, high-falutin’.

49. “We should go to the petting zoo, I hear they have cattle now,” Tom said, compatible with me.

50. “After Kant’s death, he left his old machine gun to forces plotting a military coup,” Tom said, willing that his maxim could make a general rule.

# Contra Caplan on Mental Illness

I.

Bryan Caplan has a 2006 paper arguing that economic theory casts doubt on the consensus view of psychiatric disease. He writes:

Economists recognize the benefits of specialization. Only with hestitation, then, can economists focus their attention on an unfamiliar discipline and conclude that experienced professionals have been making elementary mistakes. However inconsistent psychiatry’s main theses seem to be with basic consumer theory, one might think it foolhardy to conclude that they are wrong.

At the same time, economists also recognize not only that rentseeking is a ubiquitous force, but that most rent-seekers create and internalize public-interested justifications for their activities. It is not overreaching for economists to criticize domestic auto makers’ arguments for protectionism. The auto makers know more about the details of their own industry, but economists are better at interpreting those details. Equally importantly, economists are trained to consider the costs of a policy for everyone in society, not merely groups with the most political influence.

From a rent-seeking perspective, skepticism about psychiatry is common sense. Rent-seeking is only a side activity for the auto industry, but it lies at the core of psychiatry.

Calling someone a rent-seeker is sort of an economist’s way of telling them to die in a fire, so I feel honor-bound to respond.

As best I can tell, Caplan’s argument goes like this:

Consumer theory distinguishes between two different reasons why someone might not buy a Ferrari – budget constraints (they can’t afford one) and preferences (they don’t want one, or they want other things more). Physical diseases seem much like budget constraints – the reason a paralyzed person can’t run a marathon is because it’s beyond her abilities, simply impossible. Psychiatric diseases seem more like preferences. There’s nothing obvious stopping an alcoholic from quitting booze and there’s nothing obvious preventing someone with ADHD from sitting still and paying attention. Therefore they are best modeled as people with unusual preferences – the one with a preference for booze over normal activities like holding down a job, the other with a high dispreference for sitting still and attending classes. But lots of people have weird preferences. Therefore, psychiatric diseases should be thought of as within the broad spectrum of normal variation, rather than as analogous to physical diseases.

He compares this to the work of Thomas Szaszszszsz, who proposes that psychiatry is an inherently political enterprise that works to delegitimize people with unusual preferences. For example, until the 1970s homosexuality was considered a psychiatric disease, and now it is considered an uncommon but legitimate preference. In the past being transgender was considered a psychiatric disease, but now many people are moving toward considering it an uncommon but legitimate preference. In each case, when society thinks that a preference is gross, or anti-social, or so extreme that they can’t imagine themselves having it, they shout “Psychiatric disease!” and then they can stick anyone who offends them in mental hospitals; if the preference becomes more legitimate, they retreat and say “Guess those ones weren’t psychiatric diseases after all, but we’re still 100% sure all the other ones are”. Caplan says that instead of these constant mini-retreats we should just admit that all psychiatric diseases are unusual preferences.

He admits that he’s making his job too easy with examples like alcoholism, and that something like schizophrenia would make a harder test case. But, he asks, what is schizophrenia? Delusions and hallucinations. Delusions may be a preference to have a weird belief – for example, somebody might feel trivial and neglected, so in order to make themselves feel important they cook up a paranoid conspiracy theory where the FBI, CIA, and Freemasons are all after them because of how important they are. Wanting to believe that the government is after you is a weird preference but, Caplan insists, nevertheless still a preference. As for hallucinations, sure, schizophrenic people say they hear voices, but we all kind of hear internal voices in the sense of “hear ourselves thinking in our head”, and schizophrenics must just be people who prefer to explain those in very vivid external terms. Visual hallucinations are the same way – we all have an imagination, but some of us react to our imaginations differently than others.

Therefore, all psychiatric diseases can be conceptualized in the form of preferences. This makes them very different to physical diseases, which are budgetary limitations, and we should stop saying they’re the same thing and locking mentally ill people up and having all of these rent-seeking psychiatrists around saying they can “cure” them.

Caplan ends by noting that genetics and neurobiology cannot prove him wrong. Yes, weird preferences may be genetic, and they may be linked to weird neurobiology, but so are our normal preferences! There are genetic factors influencing schizophrenia, but there are also genetic factors influencing politics, religion, and extraversion. Yes, drugs can make you less schizophrenic, but they can also make you less extraverted.

I agree with Caplan’s last paragraph. We can’t prove him wrong with neurobiology alone. So let’s prove him wrong with philosophy, psychology, economics, and common sense.

II.

Let’s start with preferences vs. budgetary constraints.

Alice has always had problems concentrating in school. Now she’s older and she hops between a couple of different part-time jobs. She frequently calls in sick because she feels like she doesn’t have enough energy to go into work that day, and when she does work her mind isn’t really on her projects. When she gets home, she mostly just lies in bed and sleeps. She goes to a psychiatrist who diagnoses her with ADHD and depression.

Bob is a high-powered corporate executive who rose to become Vice-President of his big Fortune 500 company. When he gets home after working 14 hour days, he trains toward his dream of running the Boston Marathon. Alas, this week Bob has the flu. He finds that he’s really tired all the time, and he usually feels exhausted at work and goes home after lunch; when he stays, he finds that his mind just can’t concentrate on what he’s doing. Yesterday he stayed home from work entirely because he didn’t feel like he had the energy. And when he gets home, instead of doing his customary 16 mile run he just lies in bed all day. His doctor tells him that he has the flu and is expected to recover soon.

At least for this week Alice and Bob are pretty similar. They’d both like to be able to work long hours, concentrate hard, and stay active after work. Instead they’re both working short hours, calling in sick, failing to concentrate, and lying in bed all day.

But for some reason, Bryan calls Alice’s problem “different preferences” and Bob’s problem “budgetary constraints”, even though they’re presenting exactly the same way! It doesn’t look like he’s “diagnosing” which side of the consumer theory dichotomy they’re on by their symptoms, but rather by his assumptions about the causes.

And his assumptions about the causes may be wrong. Bob’s issues are probably caused by what we call “sickness behavior”, a chemical defense in which the immune system notices an infection and releases cytokines telling your brain to avoid action and conserve energy in order to help with recovery. But one of the theories of depression I have found most plausible is that it’s a malfunctioning of sickness behavior – you’re not necessarily really sick, but your immune system releases its “stop acting and lie in bed all day so we can recover” chemicals anyway. If flu and depression have the same proximal cause, and the same effects on your life, where does Bryan draw the budget/preferences line?

For that matter, does Bryan ever get tired? I mean, suppose he is up very late one night and then has to go to work on only an hour of sleep. If he’s like the rest of us, he probably does a terrible job, can’t concentrate, and maybe rushes through things to get home early so he can catch a nap. Is this a budgetary constraint, or different preferences? In one sense it seems budgetary – he is lacking a resource (sleep? mental energy?) that would allow him to do a good job if he had it. In another sense it is clearly preferential – he places much less value in working hard and much more value in rushing home to get a nap.

Either way, this seems like a fruitful way to think about conditions like ADHD. Someone with ADHD, like someone who’s working on an hour of sleep, finds themselves miserable and unable to focus. If we call this a budgetary constraint, Bryan’s whole argument comes tumbling down. But if we call it a preference, then it’s a very strange type of preference, one where the usual method of “oh, great, you’re doing what you prefer!” is entirely the wrong approach.

In the case of the sleep-deprived person, their “new set of preferences” seems like a malign and unpleasant condition inflicted on them by an external source – namely, their sleeplessness. It’s certainly not “what they prefer” in the sense of “oh, great, he’s doing what he prefers, now he’s self-actualized and all is right with the world.” Instead, we admit that it’s a problem, they admit that it’s a problem, and we prescribe a biological cure – more sleep.

Or here’s another example. Suppose while you are asleep I inject you with a little machine that constantly releases interferon into your bloodstream – interferon being a hepatitis medication notorious for causing deep depression as a side effect. You become depressed, and your preferences change from “work and spend time with my friends” to “lie in bed all day”. But a helpful wizard gives you a powerful antidote. If you take the antidote every day, the interferon is rendered harmless and you are as active as ever. Unfortunately, you run out of antidote, and you lie in bed all day.

So: is your lying in bed all day a preference, or a budgetary constraint caused by shortage of antidote?

My answer: dichotomies sometimes break down outside a certain scope. I enjoy reading Bryan’s posts about immigration, which often compare immigrants to natives along some axis. Immigrant vs. native is a useful dichotomy for a lot of purposes. But: were the guys on the Mayflower immigrants or natives? Were slaves abducted from Africa to work on plantations immigrants or natives? Were the couple thousand residents of California who went to bed as Mexicans the night before the Treaty of Guadalupe Hidalgo and woke up as Americans the next morning immigrants or natives? What about migratory birds? The number three? The Devil?

I propose that the preference/budget distinction is a bad way of dealing with anything more complicated than which brand of shampoo to buy. We intuitively talk about our choices as if there were some kind of “mental energy” that allows one to pursue difficult preferences, and I discuss some ways this confuses our intuitive notion of budgeting in Parts II and III here. You don’t have to accept any particular framing of this, but to sweep the entire problem under the rug is to ignore reality because you’re trying to squeeze all of human experience into a theory about shampoo.

III.

If there’s not a lot of difference between preferences and budgetary constraints, what does that say about the relevance of Thomas Szasz?

(By relevance, I mean relevance to the modern day; he wrote in the 1960s and what he wrote was very possibly more true then than it is now. Possibly played a big part in making the things he wrote about less true, and should be celebrated for this. But I’ll concentrate on the present.)

Szasz and Caplan both says that mental illnesses are attempts to stigmatize those with unusual preferences. I say that mental illnesses can reflect people’s genuine worries about a-thing-sort-of-like-a-budgetary-constraint afflicting them. Which of us is right?

Well, consider that about 95% of people who go to an outpatient psychiatrist do so of their own free choice. This is certainly the case with my own patients. They are people who have gotten tired with the constraints that mental illnesses put on their lives, come in and say “Doctor, please help me”, and I try to help them achieve whatever goals they have for themselves.

About 50% of people who go to inpatient psychiatric facilities also go of their own free choice. The rest, assuming everyone’s following the legal system and the appropriate ethical guidelines, are people who are “dangers to themselves or others”. I admit, it takes a controversial value judgment to decide people shouldn’t be allowed to be dangerous toward themselves – though I think in some cases that judgment can be justified. And I admit that “danger to others” can sometimes be stretched to the point where if a psychiatrist wants to commit someone they can probably make up a justification. But these implementation problems are a heck of a long way from Caplan and Szasz’s theory of “psychiatry is just a project about finding weird people and locking them up.”

The psychiatric profession will never live down the thing about homosexuality; I fully expect that in 5000 AD someone will still be complaining that we can’t stigmatize entities infected with superintelligent self-replicating memetic viruses, because DSM-II listed homosexuality as a psychiatric disease. But there’s still a chance to rebut the thing about transgender, so let me quote from the APA website’s discussion of the topic:

The [new DSM criteria] underscore that being transgender is not a disorder in itself: Treatment only is considered for transgender people who experience gender dysphoria — a feeling of intense distress that one’s body is not consistent with the gender he or she feels they are, explains Walter Bockting, PhD, a clinical psychologist and co-director of the LGBT Health Initiative at Columbia University Medical Center.

In other words, the decision about whether transgender people need psychiatric help is left up to – transgender people. If they don’t want it, they don’t have to have it. If they do want it, the option is open to them and their condition is recognized as a legitimate reason to seek help that insurance companies will support. I myself have treated exactly one patient for gender dysphoria. She was so depressed about her gender that she was considering suicide. I gave her some antidepressants, some supportive therapy, and some information about local support groups and sex-change professionals. Then I billed her insurance company for gender dysphoria treatment and got a check. Truly everyone involved is Worse Than Hitler.

IV.

Caplan admits that some mentally ill people seek help voluntarily and are among the most vocal proponents of the “real disease” theory. In order to shoehorn this into his preference-budget dichotomy, he theorizes that this is an attempt at deception. For example, alcoholics’ insistence that they cannot resist drinking alcohol is deceptive:

From an economic point of view, however, what is so puzzling about a person who prefers consuming alcohol to career success or family stability? Life is full of trade-offs. The fact that most of us would make a different choice is hardly evidence of irrationality. Neither is the fact that few alcoholics will admit their priorities; expressing regret and a desire to change is an excellent way to deflect social and legal sanctions.

But in order to fully explain alcoholic behavior, we have to take this theory exceptionally far. Consider a typical alcoholic drinks for several years, then “hits bottom”, goes sober, and joins Alcoholics Anonymous. He attends AA meetings three times a week for three years, then has a really bad day and binges on alcohol. Afterwards he is so embarrassed that he attempts suicide, but is rushed to the hospital and resuscitated successfully. After that he goes back to his AA meetings.

Does this man have a preference for going to AA meetings three times a week for several years then getting really drunk then attempting suicide? That’s a weird preference to have. Does he have a preference to drink, and in order to be socially acceptable he ‘covers up’ his one episode of binge drinking by years of AA meetings and a serious suicide attempt which he secretly knows will fail? That is a pretty disproportionately big web of lies, especially when probably no one would blame him for binge drinking one night one time.

If we’re willing to be this paranoid, we can basically prove or disprove anything. Bryan Caplan says he’s a libertarian, but my 9th grade Civics textbook says there are only two political parties, Democrats and Republicans. If Bryan says he’s in a third, he must just be trying to “deflect social and legal sanctions”. Maybe he’s secretly a Republican, but he wants to fit in to academic culture, so he says all of this stuff about “libertarianism” as a cover. His work writing hundreds of essays and some pretty decent books supporting his libertarian viewpoint are to maintain the credibility of his signal and throw us off the trail. Any donations he may have made to libertarian causes are the same…

…or we can be skeptical of textbooks that try to reduce things to simple dichotomies, whether that’s Democrat/Republican or preference/budget.

Caplan sort of flirts with admitting this:

Cooter and Ulen probably speak for many economists when they deny that the preferences of the severely mentally ill are well-ordered. But in fact, not only do individuals with mental disorders typically have transitive preferences; they usually have more definite and predictable orderings than the average person…it is also implausible to interpret most mental illness using a ‘hyperbolic discounting’ or ‘multiple selves’ model. These might fit a moderate drug user who says he ‘wants to quit’…but they do not fit the hard-core drug addict whose only wish is to be left alone to pursue his habit. The same holds for most serious mental disorders: they are considered serious in large part because the affected individual continues to pursue the same objectionable behavior over time with no desire to change.

But if we take that middle part seriously he is ceding me 99.9% of the ground without remarking on it. Most people with mental disorders and substance abuse disorders wants to get rid of their disorder or at least alleviate the worst parts of it. If you are willing to accept complicated “multiple selves” models for those, then that is what you should be using to model mental disorders, not the simple consumer price theory.

And the others? The alcoholic who says “Yup, I’m drinking myself to death and you can’t stop me?” I agree that it is in some sense rational. It is rational because that person has so many problems that drinking alcohol becomes more pleasant than dealing with them. Often, these problems are related to psychiatric issues – for example, many people with PTSD become alcoholics because alcohol helps them briefly forget their traumatic memories. There are many people who say they don’t want help with their drinking problem because they expect “help” to mean “take away the alcohol but give them nothing in exchange”. If “help” meant “replace the alcohol with some healthier coping mechanism that works just as well”, many of these people would take it in a heartbeat. I realize this doesn’t quite disprove Caplan’s thesis for this relatively small group of alcoholics, but I think it’s important to remember that “preference” is different from “they’re doing what they want and all is well”.

V.

Finally, Caplan moves into a discussion of schizophrenia. He says that hallucinations might just be people hearing their normal inner voice and seeing their usual inner imagination, but they choose to describe it differently:

Szasz similarly maintains that many alleged hallucinations are only eccentric descriptions of ordinary experience. To take the most common form, psychiatrists routinely equate ‘hearing voices’ with auditory hallucination. But when a person feels guilty, we often say that he hears the voice of conscience…to take a stronger case, the DSM treats ‘a voice keeping up a running commentary on the person’s behavior or thoughts’ as an exceptionally serious symptom. But this describes any person deliberating between major life options over an extended period of time. While these examples might seem to stretch the meaning of ‘hallucination’, it is the DSM that explicitly fails to distinguish whether the source of the voices is perceived as being inside or outside of the head.’

This makes sense, which is why every psychiatrist for the past century has specifically asked patients whether that’s what’s happening before diagnosing them with anything. Any time a patient reports a hallucination to me, the first question I ask is whether they’re just embellishing on hearing an inner voice, or whether they actually heard an external voice clearly and distinctly the way they are hearing me talk to them right now. Sometimes they did just hear an inner voice – this is especially common in OCD obsessions – but other times they tell me that no, it was definitely an external voice, totally different from their normal internal voice. Sometimes they thought at first it was a normal non-hallucinatory voice talking to them, and they got up to try to figure out who it was before they realized no one was around and it had to have been a hallucination.

This should not be surprising to anyone who has ever taken drugs, heard from people who took drugs, or been vaguely aware of the existence of drugs. Drugs can cause vivid, realistic hallucinations. Caplan says he doesn’t want to talk about neurobiology, and that’s all nice and well, but drugs provide a pretty good neurobiological proof of concept. LSD, which is infamous for its hallucinations, is a 5-HT2A agonist. You can treat schizophrenic hallucinations with Seroquel, which is a 5HT2A antagonist; placebo Seroquel doesn’t work nearly as well. Coincidence? I feel like at this point we’re getting into paranoid are-we-sure-anyone-is-a-libertarian territory again.

He then goes on to say that delusions might just be a preference to believe something, rather than actually believing it. There’s a lot of epistemological complexity here – can we believe something just because we prefer to believe it? If someone offered me $1 million to believe that Greenland is in the southern hemisphere, could I do it? I think not, but I think Caplan understands this and is accusing delusional people of just playing a sort of LARP where they act as if they’re in a much more interesting universe full of FBI agents and secret radios and the Devil. In favor of this, he describes how psychotic people can sometimes adjust their thinking and actions when they have incentive to do so. For example, he talks about some psychiatric inpatients denying their delusional beliefs in order to avoid electroconvulsive therapy.

I don’t think the ability of psychiatric inpatients to hide their condition in response to incentives changes things much. I firmly and genuinely believe that Greenland is in the northern hemisphere, but if someone threatened to give me old-timey scary electroconvulsive therapy for believing this, I would tell them it was however far south they wanted it to be. This doesn’t mean my belief about Greenland is insincere, it just means I can think strategically. That even very deeply mentally ill people can think strategically can sometimes be surprising, but no one who has worked with them would deny it can be true.

On the other hand, some of them can’t think strategically. I remember one patient who was very angry at being involuntarily kept in the hospital who would come up to me every day and start screaming at me that if I didn’t let him out, his friends in the highest level of the government were going to revoke my medical license – later this escalated to “kill me”. Spoiler – this is a really bad way to get out of a psychiatric institution. I even told him this was a really bad way to get out of the psychiatric institution and he’d be making his case better by just leaving me alone. He kept finding me and screaming threats about his friends in the government anyway. I can’t tell you for sure what the difference is between people who think strategically and who don’t think strategically – it’s certainly not as simple as “people with illness X are strategic, but people with illness Y aren’t”, but it is certainly a pretty obvious dichotomy.

This patient’s story continues – I put him on antipsychotics, and after two weeks he said he was feeling much better, no longer talked about his friends in the government, and actually thanked me for treating him. I discharged him and as far as I know he’s still taking those medications. If schizophrenia was a preference, this would be strange: he prefers to be schizophrenic, he knows that taking medications will make him less schizophrenic, but he keeps taking them anyway! Since many people like him become schizophrenic, keep taking their antipsychotic medication, and then become better – leaving them much as they were before they became schizophrenic – Caplan’s theory can only theorize that they have a base-level preference for being on antipsychotic drugs. This is a terrible preference to have – such drugs often have bad side effects and make you feel miserable. Surely it makes more sense to believe they have a problem which they don’t like and which the antipsychotics successfully treat?

(yes, there are also many schizophrenic people who don’t voluntarily take their medication. But there are also many people with high blood pressure who don’t take their medication, and antipsychotics are way less pleasant than antihypertensives)

One more thing. Although hallucinations and delusions are the flashiest symptoms of schizophrenia, they are by no means the only ones or even the most important. Many schizophrenics have what’s called “formal thought disorder”, which means their thoughts go in weird directions. A classic example is the tangential person, who will get so distracted they can’t finish a thought. “Tell me how the medication is working?” “Well, I took my medication this morning, after waking up, because I had a bad dream last night, I can’t remember exactly what it was about, I think there was a dog in it, my favorite kind of dog is a Labrador Retriever, I think they’re from Canada, I was in Canada once, it was really cold.” Another kind is the clang, where they connect thoughts based on sound rather than meaning: “I took my medication this morning, it was a warning, a warning of doom, coming at noon with the moon.” This is, as far as I can tell, not something that schizophrenics can successfully “tone down” when asked to do so, based on informal experiments where I ask schizophrenic people to speak normally and tell them that I am more likely to let them out of the hospital if they can form a coherent sentence. This avoids the strategic issues involved in “covering up” hallucinations. Sure, you can always have hallucinations but say you aren’t, but it’s really hard to fake not having a formal thought disorder if you have one, and indeed when a schizophrenic person has a formal thought disorder it’s there to stay until they are treated.

So as far as I can tell schizophrenia includes real hallucinations and delusions, real formal thought issues that the patient usually cannot control, and often the patient is unhappy with it and will willingly take medication to get rid of it. Combined with the neurobiological evidence, the genetic evidence, and the pharmacological evidence, I don’t think calling it “different preferences” is remotely viable.

The Caplan paper is from 2006. I don’t know if he still believes it. And I don’t know if anyone else holds this particular view. But I still meet the occasional Zsazsaian, and general feelings that psychiatric illness isn’t “real illness” are still common. I don’t think it’s a very tenable position and I don’t think this paper does much to support it.

EDIT: I previously wrote in more detail about the difference between “disease” and “normal variation” here.

# Against Against Autism Cures

[Content warning: autism, disability, psychiatry, abuse]

I.

Vox: We’ve called autism a disease for decades. We were wrong.

I have mixed feelings about this. On the one hand, I know and like many people in the autism rights movement. They defend autistic people’s right to avoid psychiatric care if they don’t want it, and this fits well with my own position that (outside of emergency situations) everyone should get to choose whether or not they want psychiatric care. They also do good work exposing abuse of autistic people and ideas that promote such abuse, and calling out organizations that claim to speak for autistic people but don’t do a very good job. And they provide great resources that help autistic people and their friends and families. Ninety-nine percent of what they do is unquestionably great.

Yet somehow, whenever I hear about them in the media, the article is titled either “Autism Is Not A Disease” or “Stop Trying To Cure Autism”. I don’t know if it’s just the usual controversy-mongering or whether this really is at the center of their philosophy, but they manage to consistently emphasize the one percent of their philosophy I can’t be on board with.

All psychiatric categories are a mishmash of unlike things crammed together under a single name. Depression ranges from people who put on a normal facade but feel empty inside, all the way to people who are totally catatonic and can’t move or speak. Schizophrenia ranges from people who are totally okay as long as they take their medication, all the way to people who talk in “word salad” because their thoughts are so malformed that they can’t even make complete sentences. But even among diagnoses like these, autism takes the cake in terms of heterogeneity.

I kind of a have a front-row seat here. On the one hand, about half my friends, my girlfriend, and my ex-girlfriend all identify as autistic. For that matter, people keep trying to tell me I’m autistic. When people say “autistic” in cases like this, they mean “introverted, likes math and trains, some unusual sensory sensitivities, and makes cute hand movements when they get excited.”

On the other hand, I work as a psychiatrist and some of my patients are autistic. Many of these patients are nonverbal. Many of them are violent. Many of them scream all the time. Some of them seem to live their entire lives as one big effort to kill or maim themselves which is constantly being thwarted by their caretakers and doctors. I particularly remember one patient who was so desperate to scratch her own face – not in a ‘scratch an itch’ way, but in a ‘I hate myself and want to die’ way – that she had to be kept constantly restrained, and each attempt to take her out of restraints for something as basic as going to the bathroom ended with her attacking the nurse involved. This was one of the worse patients, but by no means unique. A year or so ago, after a particularly bad week when two different nurses had to go to the emergency room, the charge nurse told me in no uncertain terms that the nursing staff was burned out and I was banned from accepting any more autistic patients. This is a nurse who treats homicidal psychopaths and severely psychotic people every day with a smile on her face. When she says “autistic”, it seems worlds apart from the “autistic” that means “good at math and makes cute hand flap motions”. When a mental health professional says “autistic”, the image that comes to mind is someone restrained in a hospital bed, screaming.

The Vox article doesn’t deny any of this. Matthews admits that “people have a image of autistic people as these completely nonverbal children banging their heads against the wall” and his interviewee Silberman describes how autistic children in institutions, when left alone, would “end up chewing through their own finger”, which pretty much 100% matches my expectations of what my autistic patients would do if the nurses weren’t there to watch them.

But as per Matthews and Silberman, this is just a consequence of the maltreatment these children receive in institutions:

What society thought of as the natural course of autism was actually a very skewed view of what happened to autistic people when they were put in institutions. For decades, the recommended course of treatment for autism was institutionalization.

Parents were routinely told they should put their child in an institution, quietly remove their photographs from the family albums, never speak of them again, and enlist in decades-long courses of psychoanalysis to think about why they were motivated to wound the developing psyches of their children.

When children were put in institutions for the rest of their lives, it wasn’t like they were put in specialized autism wards. There was no such thing, with very few exceptions. They were mostly put on psych wards for adult psychotics. Oliver Sacks worked on such a ward, Ward 23 at Bronx psychiatric, in the 1960s. He told me that some of the children and young adults would be put in straight jackets and isolation rooms to sit in their own waste for weeks on end. The children and young adults became self-injurious, which is not a surprise at all. If you treat people brutally, they’ll react in extreme fashion.

Be sure to read the rest of the article for more horror stories about the ways autistic children were treated; this sort of stuff was all too common, slightly-toned-down versions of it are still too common today, and this is why I agree with ninety-nine percent of everything the autism movement says.

But then they start talking about how we don’t need a cure.

II.

The popular literature about autism tends to fall into the genre of “Doctors hate her! Area mom cured her child’s autism with This One Weird Trick!” Common One-Weird-Tricks include gluten-free diets, casein-free diets, massive multivitamin doses, and whatever else the cutting edge of quackery can dream up. The autism rights people are rightly suspicious of this entire category.

But I worry they have their own One-Weird-Trick: treating autistic people decently.

You should treat autistic people decently because it’s the right thing to do. But it is not One-Weird-Trick. Avoiding abusive treatment will prevent things from being worse than they have to be, but that’s all. It will not turn severely disabled people into independent, happy-go-lucky angels. Parents, caretakers, and doctors can do everything right, move heaven and earth to accomodate an autistic person’s every need and limitation – and they will still suffer.

Some of this is purely biological. Thirty percent of autistic people have comorbid epilepsy, often very severe. Over half of autistics are cognitively disabled. Autistics have three times the risk of Tourette’s Syndrome, five times the risk of cerebral palsy, about a hundred times the risk of tuberous sclerosis, and various balance and coordination disorders, plus an increased rate of other psychiatric disorders like bipolar and schizophrenia. There are treatments for these conditions, both pharmacological and otherwise, but they come with their own set of side-effects and difficulties and none of them are 100% effective.

And some is behavioral – but as far as I can tell in no way limited to maltreated or institutionalized autistic people. Half of autistic children self-injure, and more than half of autistic children and adolescents are physically aggressive. Autistic children are twenty-eight times more likely to be suicidal than other children. Three-quarters have eating problems ranging from “picky eater” to “will not eat food, good luck doing something about this”. About two-thirds have “sleep disorders”, which is sometimes a euphemism for “wakes up screaming in the middle of the night and will not stop”. As best I can tell, all these studies were done on non-institutionalized autistic people who were generally well-treated and still living with their parents.

And yes, institutionalization adds a whole lot of extra suffering to the mix. But even here, I find Matthews’ narrative overly simplistic. He talks of a world where random doctors swoop in and trick parents into sending their children to institutions out of pure prejudice and stigma. Anyone who’s ever worked with these families has seen something very different. These parents aren’t poor deluded rubes who have been tricked by stigmatizing doctors. They’re well-educated, deeply committed to their children – and desperate. They’ve spent years trying to raise kids who were violent, self-injurious, locked in a sensory hell without the ability to explain their problems verbally, and maybe having seizures all the time to boot. Their decision to institutionalize is a reluctant concession to this reality. I do not feel the slightest bit of qualification to pass judgment upon these parents and I invite anybody who does to spend a few moments talking to them. Unless you can give these parents a better option – and trust me, they’ve looked – institutionalization isn’t a moralistic tale of prejudice and stigma in the medical system. It’s just a few more drops of misery added to the vast morass of suffering that is severe autism.

Matthews and Silberman speak dismissively of Leo Kanner’s view that autism was caused by bad parenting, but I worry we’re headed right back in that direction. Start telling people that the only reason autistic people have all these problems is because of institutional abuse, and someone’s going to ask “But what about all these autistic people living with their parents who also head-bang/attack people/try to kill themselves?” Then you can either admit that sometimes autism just sucks no matter how good an environment you’re in, or blame the parents for not making the environment good enough. And once you’ve made “autism isn’t a disease and nobody needs a cure!” into your rallying cry, it’s going to be hard to choose that first option.

III.

If we can’t make all autistic people independent and well-adjusted with One Weird Trick, then we have to consider how real autistic people actually turn out. The numbers aren’t good.

Outcomes In Adults With Autism Spectrum Disorders: A Historical Perspective reviews all of the autism outcome studies of the past fifty-odd years. Most of it is nitpicking different people’s definition of “poor” outcome versus “very poor” outcome, but let me try to extract the easily quantifiable bits1.

Six studies have assessed what percent of adult autistics have a job – they find 22%, 21%, 31%, 4%, 4%, and 4%. The two that found rates in the twenties limited themselves to high-IQ autistics and so are unrepresentative.

Four studies assessed institutionalization rates among adult autistics, although these “institutions” form a very heterogenous category from homey group houses to super-intense locked hospitals. These studies find 35%, 43%, 48%, and 53% of adult autistics to be instutitionalized.

A few studies looked at other outcomes. Two investigated what percent of adult autistics still lived with their parents. Both estimated about 50%. This is in addition to the 40% or so who are institutionalized, so only about 10% of adult autistics live independently.

One study investigated how many autistics have at least one friend and found it was just under 50%.

I cannot find any studies on adults with autism per se, but adults with Aspergers (recently collapsed into the autism diagnosis) are ten times more likely to be suicidal than other adults.

I realize this seems extreme2, but I think it really puts into perspective the difference between the conventional “shy person who likes trains” view of autism, and what psychiatrists and scientists really mean when they talk about an autism diagnosis. The happy, independent autistic people whom most of us know and whose stories get told in the media are four to ten percent of the autistic population. What about the other extreme, the forty percent who are institutionalized?

I hate to have to criticize institutions – an umbrella term I’m using to cover group homes, locked facilities, nursing homes, hospitals, etc. Many are run by amazing and caring people who are doing thankless work on shoestring budgets. I’m humbled by the patience and compassion I’ve seen in their staff of nurses, techs, and other caretakers, and I can’t judge them nor claim that I could do their job for one minute.

That having been said, a lot of institutions are kind of hellish.

I’ve seen institutionalized patients who were sexually abused, physically abused, or just neglected and left to wander out into the street. Lest you dismiss this as crazy people making things up, some of the allegations later got confirmed by police investigations.

Other times it doesn’t rise to the level of anything criminal, just the usual petty tyranny you expect any time somebody gets to control somebody else’s life. I got consulted for “medical management” of an autistic man who had smashed a bunch of holes into the wall of his group home. He needed his mp3 player to control his noise sensitivities, but the staff had taken it away as punishment for breaking some rule or other. He decided a proportionate response was to smash several holes in the wall. To his credit, it worked; my diagnosis was “give him back his fricking mp3 player, you morons”.

But even when the institutions are well-run, non-abusive, and dare-I-say-it-even-nice, the whole structure just makes a perfectly bad fit with autistic people. Remember, autistics are known for intense sensory sensitivities and pickiness about their environment. Take someone who can remain stable as long as there are no unexpected loud noises, and make him share a room with a guy who screams at the top of his lungs every couple of minutes for no reason. Or take someone who will eat about 1% of foods, and only if they’re perfectly prepared, and then stick him in a hospital where the catering service lets him choose from a menu of two meals, and they’re always out of one of them. Or take somebody who freaks out about the feeling of different textiles, and tell her she has to sleep in a mass-produced hospital bed with exactly the same sheets as every other mass-produced hospital bed.

I have some very minor sensory sensitivities, and they drive me nuts. But at least they only metaphorically drive me nuts, That’s because I’m an independent middle-class person who can throw money at problems to make them go away. I have no tolerance for stray noise, so I pay a little more than I should for housing and live in a quiet area outside town. I can’t stand constricting or scratchy clothing, so I only wear loose-fitting clothes, extend the collars, and cut off all the tags. It works pretty well – but only because I have a lot of control over my life. Put me in an institution where other people manage everything about my daily routine, and my life would fall apart. And my sensitivities are a fraction of a fraction of what real autistic people suffer.

The hospital I work in really does try hard to make things more tolerable for our autistic patients. It’s never enough. Fix all of the contigent things, and you just bump up against the fact that humans were not designed to live in psychiatric hospitals, autistic humans least of all.

I have met many well-intentioned people who believe that institutions are only bad because of stigma against the mentally ill, or insufficient budgeting, or ignorance of people’s true needs. These people should visit a nursing home someday. The people there don’t have some kind of exotic stigmatized hard-to-understand condition. They’re just old. Yet nursing homes and other institutions for the elderly have every bit as many problems as the institutions for autistic people.

I don’t think that mental health institutions will get better anytime soon, because I would expect the average person to be a lot more concerned with nursing homes – where their grandparents live! Where they themselves will end up one day! – than they are with mental health institutions. If we can’t even get our act together on that one, what hope do we have for the harder problem? Having good involuntary communal living institutions is just plain beyond us as a civilization at this point. It has nothing to do with stigma or prejudice. Even if everybody loved autistic people exactly as much as they love their own grandmother, the best we could hope for is institutions that treat autistic people as well as they treat grandmothers. Which is to say, abysmally.

V.

Dylan Matthews says that autism “is not a disease”, joining writers from TIME, The Guardian, The Irish Examiner, various blogs, et cetera. I would hate to contradict such an array of eminent voices.

So let’s taboo whether something is a “disease” or not. Let’s talk about suffering.

Autistic people suffer. They suffer because of their sensory sensitivities. They suffer because of self-injury. They suffer because they’re in institutions that restrain them or abuse them or just don’t let them have mp3 players. Even if none of those things happened at all, they would still suffer because of epilepsy and cerebral palsy and tuberous sclerosis. A worryingly high percent of the autistic people I encounter tend to be screaming, beating their heads against things, attacking nurses, or chewing off their own body parts. Once you’re trying to chew off your own body parts, I feel like the question “But is it really a disease or not?” sort of loses its oomph.

My moral philosophy doesn’t contain a term for “is this a disease or not?”, but it definitely contains a term for suffering. If you’re a good person, you try to alleviate or prevent suffering. Accomodating and supporting autistic people alleviates some amount of the suffering associated with autism. Curing it alleviates all of that suffering.

And remember – society is fixed but biology is mutable. Which do you think is more likely? That soon biologists will discover a molecular cure for autism? Or that soon politicians will discover a cure for the systemic issues that cause poor people who can’t stand up for themselves to be maltreated and abused? The biologists seem to have about a ten million times better track record for this sort of thing. And if you don’t expect the politicians to create a brave new world where no disability ever remains unaccomodated, then stopping the biologists just means that the status quo will go on forever.

Faced with the choice of seeing the flood of human misery that I have to deal with every day continue mostly unabated, or having a pill that provides a quick fix to said flood, I wish with all my heart for the latter. Sure, this should not be pursued at the cost of supplying what accomodations to existing autistic people we can, any more than blue sky cure-for-cancer research should be pursued at the cost of treating current cancer patients, but it’s right and proper to want it, to think it would immensely improve thousands of people’s lives.

Would something be lost if autism were banished from the world? Probably. Autistic people have a unique way of looking at things that lets them solve problems differently from everyone else, and we all benefit from that insight. On the other hand, everyone always gives the same example of this: Temple Grandin. Temple Grandin is pretty great. But I am not sure that her existence alone justifies all of the institutionalizations and seizures and head-banging and everything else.

Imagine if a demon offered civilization the following deal: “One in every hundred of your children will be born different. They will feel ordinary sensations as exquisite tortures. Many will never learn to speak; most will never work or have friends or live independently. More than half will consider suicide. Forty percent will be institutionalized, then ceaselessly tyrannized and abused until they die. In exchange, your slaughterhouses will be significantly more efficient.”

I feel like Screwtape would facepalm, then force him into remedial Not-Sounding-Like-An-Obvious-Demon classes.

VI.

My medical ethics have always said that outside emergencies, people who want psychiatric help should be able to get it, and people who don’t want psychiatric help should be able to refuse it.

So when autistic people say they don’t want cures forced on them, I say – fine. If you’re happy with your autism, and it’s not hurting anyone else, keep it.

But when they say we should stop all research into cures so that nobody else can have one either even if they want it, that’s a different story. When they say that, well, then I’m not the one dictating to other people what neurotypes they are or aren’t allowed to have.

When I see an autistic guy in a hospital room screaming and trying to chew their fingers off, I feel like that guy would probably want a cure for autism. Granted, that guy can’t always talk and tell me what he does or doesn’t want. But he certainly doesn’t seem happy with the status quo. And some autistic people, even some very high-functioning autistic people, have told me straight out that they want to be cured. Who the heck are we to tell them their desires are wrong?

More controversially, I think caretakers who wish there was an autism cure that could relieve them from the responsibility of caretaking have avalid wish. Nobody is entitled to another person’s life. A schizophrenic man has the right to stay off antipsychotics, but his wife has the right to make an ultimatum: “I can’t deal with you being schizophrenic, either you take your meds or I’m leaving.” Likewise, an autistic person has the right to stay autistic, but a caretaker has the right to say “This caretaking is too much for me, either take the cure or find somebody else.”3

Even more controversially, I think parents have the right to decide if they want to have an autistic child or not. I am generally pro-choice. As best I can tell fetuses have less personhood than cows, and I had a cheeseburger for dinner last night. If someone wants to abort a fetus because their pregnancy was an accident, because they don’t feel ready to have a child, because there’s some kind of problem in the family – all of those seem to me like a decision that a mother is perfectly within her rights to make, becauses fetuses are not very important moral agents. And if instead of just “I don’t want a child”, an expecting mother is worried because she doesn’t think she will have the resources or compassion or strength of will to take care of an autistic child, that is also her decision to make. Also, I feel like if your whole argument is that autism only goes badly when autistic children are mistreated, maybe you shouldn’t be simultaneously demanding that women who really don’t want autistic children and don’t believe they can take care of them should be forced to have them anyway. On the other hand, if a mother wanted an autistic child, and she was somehow sure that she could support that child and help him or her flourish, I would be okay with her aborting however many neurotypical fetuses it took.

But even more controversially, absent such certainty that your child will flourish I think if some kind of genetic-engineering autism-cure existed, parents would have a moral obligation to use it. Consider an analogy to fetal alcohol syndrome. People with fetal alcohol syndrome seem less happy, less able to achieve their goals, and more likely to suffer than people without the condition. Therefore, we have a very strong social norm that you shouldn’t drink too much during pregnancy. But people with autism also seem less happy, less able to acheive their goals, and more likely to suffer – in fact, autism shares a lot of symptoms with fetal alcohol syndrome, like seizures and intellectual disability. The moral imperative to take the hypothetical-genetic-autism-cure during pregnancy seems as clear as the moral imperative not to drink alcohol, and for the same reason4.

I don’t mean to propose a zero-sum game here. All sorts of breakthroughs and possibilities could give both sides everything they want. Maybe there is some way to alleviate some of the more distressing symptoms of autism while leaving the deeper and more cognitive parts intact. I don’t think anyone wants, as a terminal value, for people to be intellectually disabled or have more seizures. If Dylan Matthews were to list all of the things he likes about autism, and I were to list all of the things I don’t like, maybe there wouldn’t be anything on both lists. And maybe the bad things are controlled by separate genes from the good things, so that we could turn off one set of genes and not the other.

And I’m still in favor of basic income. If everyone gets a basic income, autistic people who want to spend it on getting the care and support they need to remain autistic could do so without costs to anybody else. I do hope that there will be only a very short stretch of time between the development of technology that can genetically engineer people at will, and solving the scarcity problems that make our desires sometimes demand sacrifices of others.

And maybe the best option is that somebody develops technology that can change your brain type during adulthood, so that everyone has the option of experiencing life as an autistic person and as a neurotypical person and seeing which is best for them. Invent something like that and the problem disappears.

And a lot depends on the genetic structure of autism. If autism is just the brain’s response to high mutational load, then trying to do most other good things will prevent autism as a side effect. On the other hand, if autism is the extreme version of a trait whose common version is “good at logical and scientific thinking”, then we have to ask ourselves whether that trait is worth messing with. As usual, everything about genetic engineering raises thorny scientific and ethical quandaries, and I can only hope we don’t drag our feet in creating the eight-foot-tall IQ 300 supermen who can solve them.

And maybe all of this is silly, because we don’t have an autism cure and we’re not even very close to one, and maybe when people talk about whether we should use the autism cure we don’t have, it’s all a metaphor for “respect my tribe” and “I demand my rights” and “here’s a good controversial lead-in to a story about how you should treat autistic people like human beings” and maybe even if I’m right about the literal moral dilemma I should support the opposite side for totally symbolic reasons.

But if, after all this, it really does comes down to a binary “cure autism/don’t cure autism” decision, I know which side I’m on.

Footnotes

1. In a way, all of these numbers are meaningless. If you define “autism” so broadly that it includes Dylan Matthews and me, maybe less than one percent of autistics are in institutions. If you define “autism” so strictly that you only count institutionalized autistics, then one hundred percent of them are. So really what these studies are saying is “We chose to define autism at a level where forty percent of autistics are in institutions”, to which one response is “And why should I care what level you chose to define autism at?” But I think these studies at least tell us two things. First, that the formal psychiatric definition of autism is much stricter than the popular one, and we combine the two categories at our peril. And second, that this former category of very severly ill autistics exists and contains many people we might not otherwise notice.

2. The employment and living-with-parent numbers don’t seem to have changed much over time; the institutionalization number has decreased over the past few decades. This probably reflects a general trend toward deinstitutionalization throughout the mental health world. So far there is little sign of autistic people doing great now that the most abusive 60s-style institutions have been curtailed. But perhaps this just hasn’t shown up yet; these are studies of adult autistics; even the youngest were born in the 1980s, and the autism world has changed a lot since then. I’d say that I will be interested to see a similar study in 2030, but I think by that time the diagnostic category will have become so confusing that nothing will be comparable to anything that came before it.

3. Even more controversially, I think the government has the right to do the same. The lifetime cost of supporting an autistic person is $1.5 million, not including productivity loss of the person themselves. Much of this is borne by the government. The average taxpayer will give about $500,000 over their lives, so it takes three non-autistic people to support each autistic person – even ignoring all other essential government services like schooling and welfare and giant nuclear missiles. Of course, helping the needy is exactly the sort of thing the government should be doing, and right now the government ought to do whatever it takes. But if there is a cure for autism and autistic people choose not to take it but still want the $1.5 million, then the government has the right to start thinking things like “we could lift a whole lot of destitute families out of poverty for $1.5 million.” Everyone has the right to choose their neurotype, but I’m not sure they have the right to make other people subsidize it.

4. Prof. Mora LeQuivalence: “Interesting point! Let me propose a similar argument that you might have a more personal stake in. Why don’t we genetically engineer away nerdiness? Everyone knows nerds are less happy than normals, at least in high school.”

Scott: “The cost-benefit calculations are totally different. Nerds are probably unhappy, but less so than autistics. And while autistics can claim Temple Grandin, nerds can claim pretty much every decent scientist and mathematician of the past few centuries, plus all the good sci-fi/fantasy writers.”

Mora: “Fine then. Keep the nerds around until they build robots that can do math and science and art. Then we won’t need them anymore and we can get rid of them.”

Scott: “Look, we can’t do all of this based on broad principles. There’s a balance between having a wide variety of human experiences versus having some of them be hellish. Few people want all individuality collapsed into a planet-sized chunk of hedonium, and few people think that a couple of expecting mothers should drink lots of booze to protect the vital diverse human experience of fetal alcohol syndrome. You’ve got to make your own choice about where to draw the line, and I draw it somewhere south of nerds and north of severe autism.”

Mora: “That’s a very interesting argument.”

Scott: “It better be, I stole it directly from God.”

# It Was You Who Made My Blue Eyes Blue

[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 at 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.”

“Waaiiiiit 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.

# Contra Simler on Prestige

su3su2u1 challenged status/signaling theories of human behavior: can they make any real-life predictions? His example was a recent medical conference that threw together three groups of people – high-status top professors, medium-status established doctors, and low-status new residents. The women in one group (female doctors + male doctors’ wives/girlfriends) were wearing conspicuous fancy jewelery. The women in the other groups weren’t. Which group had the jewelery?

His point was that status/signaling theories don’t answer this question for us with any degree of confidence. Maybe the high-status top professors wear the jewelery to signal wealth and dominance. Maybe the low-status new residents wear it aspirationally and because they need to impress. Maybe the medium-status established doctors wear it, because the residents can’t afford it and the professors countersignal that they don’t need it.

Now, in fact su3su2u1 was a no-good sneaky sneak, because the residents had all just attended a wedding that gave out the fancy jewelery as gifts and this was probably all that was going on. But his point is well-taken. Status and signaling theories are hard to use in practice. So it’s always nice when people try to do some theoretical work on them and tease them apart into their different components. This is the task Kevin Simler takes on in Social Status: Down The Rabbit Hole.

His theory (which he adopts from various psychologists and animal behaviorists) is that status separates neatly into two systems: dominance and prestige. Dominance is “respect me because I’ll kill you if you don’t.” Prestige is “Respect me because I’m awesome”. The two systems have different origins and different behavioral effects; conflate the two and you’ll end up very confused.

If you hate your boss, but you do what she says anyway because she’ll fire you if you don’t, that’s dominance. If you’re very respectful to a police officer because he has a gun and you don’t, that’s dominance too. Principals have dominance, parents have dominance, psychiatrists keeping you in a hospital against your will have dominance. Prestige is different. A rock star has prestige. He can’t hurt you. You don’t necessarily need anything from him. But you still want his autograph, want to meet him, maybe want to sleep with him. Star athletes have prestige. Actors and actresses. Good bosses who you work hard for not because you’re afraid of them but because you don’t want to let them down. Your parents, if you do what they say out of respect/love and not out of fear of punishment. Heroic leaders like George Washington (except more alive).

Having prestige can be better than being dominant. If you’re dominant, your subordinates will do exactly as much as necessary to avoid your wrath; if you’re prestigious, they may go above and beyond to help you. On the other hand, sometimes good old-fashioned dominance does the trick; your boss can ask you to drop everything and spend a week of long nights on a sudden project, but if your favorite rock star asked you to spend a week doing his taxes for him you might politely decline.

Dominance has clear animal analogies (alpha chimps, chicken pecking orders, etc), and we can pretty well guess why it evolved. The evolutionary origins of prestige are murkier, and this is the focus of Simler’s piece.

First he flirts with the theory of a guy called Henrich, who says prestige comes from a desire to learn. I admire and flatter my favorite rock star because I’m hoping I can hang out around him, some of his genius will rub off on me, and I’ll be able to play a wicked guitar riff and win a couple of Grammies myself. This theory makes no sense to me. It’s not just that there’s zero chance of Bowie teaching me, or that I might not have the talent anyway. Maybe in the environment of evolutionary adaptedness that didn’t matter so much. It’s that I don’t want to be a rock star, and if Bowie offered to train me, I’d say I wasn’t interested.

Simler doesn’t like this much either, so he moves on to the theory of two guys named Zahavi and Dessalles. I’ll quote him at length:

Unlike Henrich, whose account of prestige is unique to our species, Zahavi and Dessalles find analogues among non-human animals — most vividly, in the Arabian babbler.

The Arabian babbler is a small brown bird found in the arid brush of the Sinai Desert and (you guessed it) the Arabian Peninsula. It spends most of its life in small groups of three to 20 members. These groups lay their eggs in a communal nest and defend a small territory of trees and shrubs that provide much-needed safety from predators.

When it’s living as part of a group, a babbler does fairly well for itself. But babblers who get kicked out of a group have much bleaker prospects. These “non-territorials” are typically badgered away from other territories and forced out into the open, where they often fall prey to hawks, falcons, and other raptors. So it really pays to be part of a group. (Keep this in mind; it’ll be crucial in a moment.)

Within a group, babblers assort themselves into a linear and fairly rigid dominance hierarchy, i.e., a pecking order. When push comes to shove, adult males always dominate adult females — but mostly males compete with males and females with females. Very occasionally, an intense “all-out” fight will erupt between two babblers of adjacent rank, typically the two highest-ranked males or the two highest-ranked females. This is the babblers’ version of a Wild West showdown, as if one babbler suddenly turns to the other and says, “This town ain’t big enough for the both of us.” A showdown always results in death or permanent exile for one of the combatants.

Most of the time, however, babblers get along pretty well with each other. In fact, they spend a lot of effort actively helping one another and taking risks for the benefit of the group. They’ll often donate food to other group members, for example, or to the communal nestlings. They’ll also attack foreign babblers and predators who have intruded on the group’s territory, assuming personal risk in an effort to keep others safe. One particularly helpful activity is “guard duty,” in which one babbler stands sentinel at the top of a tree, watching for predators while the rest of the group scrounges for food. The babbler on guard duty not only foregoes food, but also assumes a greater risk of being preyed upon, e.g., by a hawk or falcon.

Helpfulness, bravery, heroism: these birds seem like regular Boy Scouts. At least on the surface.

But here’s where things take a turn for the weird. Babblers don’t just passively or occasionally offer to help each other. Instead they compete intensely for the privilege of doing so.

Unlike chickens, who compete to secure more food and better roosting sites for themselves, babblers compete to give food away and to take the worst roosting sites. Each tries to be more helpful than the next. And because it’s a competition, higher-ranked (more dominant) babblers typically win, i.e., by using their dominance to interfere with the helpful activities of lower-ranked babblers. This competition is fiercest between babblers of adjacent rank. So the alpha male, for example, is especially eager to be more helpful than the beta male, but doesn’t compete nearly as much with the gamma male. Similar dynamics occur within the female ranks.

Now: what in Darwin’s name is going on here? Why are babblers so eager to help each other?

The naive answer is that they’re simply doing what’s best for the group — because when the group succeeds, everyone ends up better off. But this kind of straightforward altruism simply isn’t found in nature.[1] It’s not game-theoretically stable, thanks to the free-rider problem. Also note that babblers actively interfere with the helpful behavior of their rivals. If their ultimate goal were the success of the group, interfering with others would be entirely counter-productive.

So the logic of natural selection compels us to ask, “What selfish motive does an individual babbler have to help others?”

The answer, in a word, is prestige. A second form of social status that lives alongside the babblers’ dominance hierarchy — a kind of “credit” reflecting the amount of good each individual has done for others. So when two babblers compete to stand guard duty, for example, they’re actually jockeying, selfishly, for prestige within the group.

And suddenly the intense competition makes sense.

But as in our species, so too in babblers: prestige means nothing without admiration. If other babblers weren’t willing to defer and pay respect to prestigious individuals, there’d be no incentive to compete for prestige.

But other babblers are willing to pay respect to prestigious individuals, in two main ways. The first is mating opportunities.[2] Babblers are constantly trying to interfere with their rivals’ mating attempts — but when a babbler has high prestige, his or her rivals interfere less. Among males, this translates to more mating opportunities; among females, it translates to earlier mating opportunities (giving one’s offspring a head start in the communal nest)

The other perk of high prestige is a reduced risk of being challenged to an all-out showdown. The higher a babbler’s prestige, the less likely its rivals are to pick a fight — even if they stand a good chance of winning.

All of which brings us, finally, to the point. Why do other babblers voluntarily defer to prestigious ones? The answer is simply(!) that babblers with lots of prestige are useful to the group, and therefore useful to keep around.[3] This is how it ends up being in the selfish interest of other babblers to defer to those with high prestige.

When a babbler is useful enough, in other words, it’s in the self-interest of others to “suck up” or pay respect to that babbler (by backing down from fights and interfering less in its mating attempts) in order to keep it happily in the group.

Bottom line: Prestige-seeking and admiration (deference) are complementary teaming instincts. They help babblers stay attached to a group, keep groupmates happy, and secure a larger share of the group’s reproductive “spoils.”

I hope this account of the babbler prestige system sounds familiar, because it’s more or less equivalent to the prestige system found in our own species; both are derived from the same Platonic form.

This is better. It sort of makes sense as an evolutionary explanation. But I think extending it from there to modern human prestige is a big stretch.

Take the rock star again. Let’s say David Bowie. When people admire Bowie, are they trying to get him to not leave the group? Is that why people scream and throw themselves at him? What would it even mean for Bowie to leave the group? If he doesn’t have enough groupies, will he defect to North Korea?

And don’t we sometimes admire people who we do want to leave the group? Suppose that for some reason I was stuck on a plane sitting next to the Koch Brothers – maybe all their private jets broke down at once. I would probably treat them in the classic way someone treats prestigious people. I’d feel really nervous striking up a conversation with them because they’re high-status and important. If I did strike up a conversation with them, I’d be really deferential and overthink everything they said. After the flight was over, I would immediately post to Twitter “I SPENT A WHOLE FLIGHT TALKING TO THE KOCH BROTHERS!” and then post the photo I’d roped them into taking with me. But none of this is because I don’t want them to leave the group. If the Koch Brothers defected to North Korea, that would be great.

And what about prestigious people who don’t bring any special talents to the group? Helen Keller, for example, can do less than most other people. We admire her not because we need to make use of her mad skillz, but because given all her handicaps it’s amazing that she can do anything at all.

We could potentially dismiss all of these by saying that evolved instincts don’t have to work in the present day. If there were no cavemen like David Bowie (probably a safe bet), then maybe our evolutionary instincts don’t apply to his case. But even in evolutionary time, admiration has a free-rider problem. Suppose that we want to make sure David Bowie stays in the West rather than North Korea, but he’ll defect unless at least three people flatter him per day. Assuming that flattering David Bowie involves some kind of cost – maybe you have to buy the t-shirt with his face on it – why should I pay the cost when there are millions of other Westerners invested in the same project? Should we be more impressed with the altruistic spirit of people who have sex with famous rock stars, seeing as they are sacrificing their bodies to the project of keeping their heroes out of Kim Jong-un’s clutches?

I think I might be straw-manning the babbler hypothesis here, so let’s skip down a few paragraphs to the next time Simler explains it:

The point is, we want to be friends, allies, and teammates with people who do good things for their friends, allies, and teammates. It’s in our self-interest to cultivate access to such people — which we do, in part, by paying them respect and granting them the perks of prestige.

More generally, however, we admire not only those who actually do good things for their teammates, but also those who show the potential to do good things, i.e., by demonstrating useful skills. The student who gets straight As from a good college, for example, is advertising her value to future employers, and her prestige makes her highly sought-after on the job market. She’ll be actively courted by hiring managers and given various perks (a better starting salary, more time to make her decision) that aren’t accorded to her less-impressive classmates.

Simler treats this as a summary of his previous point, but this is a very different theory!

The previous point was that prestigious people do good things for their community. The new point is that prestigious people do good things for their flatterers in particular. It’s a tit-for-tat relationship: show David Bowie your tits, and he gives you some tat. Money? Access to the best clubs? A copy of his latest album?

This makes sense except that it’s not the way most admiration-interactions actually work.

Forget David Bowie. Let’s talk about Justin Bieber. I see about a zillion teenage girls hanging posters of Justin Bieber in their room, fighting for the last ticket to Justin Bieber concerts, buying magazines with Justin Bieber on the cover. But the chance that Justin Bieber gives any tat for all of these tits is practically nil.

And we can’t dismiss this as a form of irrationality restricted to teenage girls. A lot of people I know geek out about Elon Musk; I’ve been to more than one party/meetup where the topic of conversation turns to how great Elon Musk is. I don’t hang posters on my wall, but if I did, they would probably have his face on them. But I don’t expect any repayment from him; I doubt he even knows about my flattery. What about all those Catholics who obsess over the Pope? What about people who obsess over J.K. Rowling or Neil Gaiman or LeBron James or Derek Jeter?

And what about me on that airplane with the Koch Brothers? Am I thinking to myself “If I ever need an entire field of science discredited, now I’ve got an in with some people who are really good at it”? What about Helen Keller? “If the world is plunged into eternal darkness, and also there’s some global super-loud hum that makes it impossible to hear anything, now I’ll have a friend who can operate regardless?” Even in evolutionary times, we should have some need to reflect on “can this person actually help me?”

I worry no one theory can completely explain prestige. It seems to me to be a combination of several different things:

1. Group signaling. The people I admire say a lot about me. If I admire Elon Musk, it means that I’m really into space, technology, and maybe the free market. If I admire the Pope, it means I’m really into Catholicism. If I admire David Bowie, it means I’m fabulous. Learning about these people, celebrating their accomplishments, and joining their Official Fan Clubs is an important method of bonding with other peopel.

2. Coattail riding. If a prestigious person becomes more prestigious, I might “look good” for having supported them “before they got big”. It suggests that I’m a good judge of character, or “hip” enough to know which acts will take off and which ones will never achieve broader appeal. Just as a fan feels good when his sports team wins the Superbowl, and a patriot feels good every time her country wins a war, so being a known Elon Musk fan means I get to feel a tiny fragment of the glory whenever Elon Musk invents a new rocket.

3. Prestige by association. Prestigious people hang out with other prestigious people. Nonprestigious people hang out with other nonprestigious people. If I have access to prestigious people, even in some boring trivial way, that makes me seem more prestigious. I think this is what’s going on with the hypothetical airplane conversation with the Koch brothers. Yes, in some sense it’s sheer coincidence that I run into them on a flight. In another sense it isn’t; at the very least, it probably means I was flying first class, and I must have had some rudimentary level of social skills to engage them in conversation. I’m signaling that I’m the sort of person who, at least when everything goes right, can shmooze with billionaires. Even if deep down people know that it was mostly a coincidence, on some gut level that’s kind of impressive.

4. Tit for tat. Yes, in some cases we will be close enough to prestigious people that we can expect rewards for our support. It’s probably easier to flatter my boss or my favorite teacher effectively than to flatter Justin Bieber or the Koch brothers, and you can reasonably expect special treatment. This is a good way of forging an alliance. If I praise my boss, she benefits from my elevation: having a nobody admire you is boring, but having a somebody admire you is both flattering practically useful. Therefore, the more I admire and support my boss, the more she is incentivized to help me become a somebody.

5. Virtuous cycles. Suppose that, for reasons 1 through 4, people want to be associated with prestigious people. Note that this is different from “associate with prestigious people” in the sense of meeting them directly; anything that gets their name linked to the prestigious person will work. In fact, suppose that specifically, there are a bunch of conservatives who are really into the Koch brothers and are jockeying for position as Koch brother fan #1. Some of these people might play the strategy of according me prestige for having met the Koch brothers as a way of better signaling their own respect for the Koch brothers to third parties. That gives me a separate incentive to seek such prestige by association.

This is still woefully incomplete, especially by “predict which of these doctors will wear jewelery”-level standards. Maybe prestige shouldn’t be treated as a single thing at all. Maybe the admiration I feel for my boss (a real person in my social circle who I interact with daily) comes from a totally different part of the brain and has totally different evolutionary origins from the admiration I feel for Elon Musk (who I expect never to meet).

But I think separating dominance from prestige is a good start. Do consider reading the full Melting Asphalt essay, as well as Simler’s follow-up thoughts.

# A Whiter Shade of Candidate

Vox says that Donald Trump practices the “politics of white insecurity”. US News says that Trump shows “the rising power of the white vote”. Salon wants to tell you “eight reasons why white America falls for demagogues like Donald Trump”. The Week says Donald Trump represents “the rise of white identity politics”. The National Journal says Trump is creating problems by “preaching to a shrinking white electorate”.

Read enough of these articles, and you might start to get the feeling that Donald Trump’s supporters are disproportionately white. You would be wrong.

Well, probably. Data are sketchy. There aren’t a lot of polls that sort their questions by race, and when they do there are sufficiently few non-white Republicans that they have trouble getting a good sample size. Nevertheless, the ones we have suggest that Donald Trump’s supporters are about as diverse as any other Republican’s and maybe moreso.

An August YouGov poll with a sample of 30 Hispanic Republicans finds Trump in the lead with that demographic, getting 28% of the vote to runner-up Ben Carson’s 19%. A Gravis poll with about 40 Hispanic Republicans finds Trump with 37% of their vote to runner-up Marco Rubio’s 20%. A Nevada poll also finds Trump leading among Hispanics in that state, though I don’t know their Hispanic Republican sample size. And finally, head-to-head matchups of Trump vs. Clinton show Trump outperforming some past Republican candidates, including Mitt Romney and George Bush, in the share of Hispanic votes he would likely receive.

This picture is confused by articles asserting either that Trump has the highest favorability ratings among Hispanics, or that Trump has the lowest favorability ratings among Hispanics. In fact, both are true! Favorability ratings allow you to rate someone favorable, unfavorable, neutral, or never-heard-of-em. Everybody has heard of Trump, and nobody is neutral about him, allowing both his favorability and his unfavorability to be sky-high (with news sources reporting whichever one of those two facts suits their narrative). Some people have done a little better work and reported his “net favorability”, or favorability-minus-unfavorability, which is very low and indeed negative. But this isn’t what matters in a real election. What matters in a real election is who people vote for. If 40% of Hispanics view him favorably, but they all vote for him, and 60% of Hispanics view him unfavorably, but split their votes among the other ten candidates, Trump has won the Hispanic vote.

As hard as it is to find good data about Hispanics, it’s even harder to investigate black Republicans. The YouGov poll that had thirty Hispanics has only five blacks; it looks like three vote for Carson, one for Rubio, and one for Trump.

Probably more useful are the head-to-head Trump vs. Hillary polls, which survey all blacks (not just Republicans). One finds Trump doing shockingly well and tripling Romney’s (admittedly miniscule, admittedly decreased by opposing Obama) level of support among black voters, but the Washington Post is skeptical and cites others with less extreme results – although even most of those show Trump doing at least as well as Romney and other historical Republicans. In terms of boots on the ground, African-American Daily Beast correspondent Barrett Pitner agrees that “Donald Trump has black supporters – really”, explaining that Trump’s “fear mongering and us-vs.-them tactics have not only created a large supporter base among conservative white Americans, but also black Americans who have been disproportionately hit by the economic downturn.”

There are too few data to say anything for sure. But all of the data that exist suggest that if the Republican primary were held today and restricted to non-whites, Trump would still win. And if Trump were the Republican nominee, he could probably count on equal or greater support from minorities as Romney or McCain before him.

In other words, the media narrative that Trump is doing some kind of special appeal-to-white-voters voodoo is unsupported by any polling data.

On the other hand, there is a candidate whom the media narrative fits like a glove. A candidate who may win primary among whites, but loses in a landslide among minorities. A candidate whose black support is almost an entire order of magnitude lower than his white support.

That candidate is Bernie Sanders.

According to the same YouGov poll mentioned above, 38% of whites support Bernie Sanders for President, compared to 37% of whites who support Hillary for President. However, only 13% of Hispanics support Sanders, compared to 63% for Hillary. And only 4% of blacks support Sanders, compared to 64% for Hillary!

A South Carolina poll from this month broadly agrees. CNN finds that the two candidates are in a statistical dead heat among whites (48-47) but that Hillary has an overwhelming advantage among blacks (84-7).

Other polls are slightly less extreme but tell the same picture. Gravis (early August) found Hillary leading comfortably among all races, but Sanders’ support among whites was still twice as high as among blacks. The Washington Post also found Sanders doing abysmally overall, but his support among blacks was super-abysmal – only 5 percent!

Suppose we measure a candidate’s “whiteness” by the ratio of their level of white support to their level of nonwhite support within their party. Donald Trump seems to be somewhere around 1.3 – 1.5. Bernie Sanders is somewhere from 3 – 10. It isn’t even close. If any candidate is “playing to the politics of white insecurity” or “preaching to the white electorate” or “harnessing the white vote”, it is he.

(though I should clarify that in a general election, Sanders would no doubt garner much higher nonwhite support than Trump just because of the D after his name. We’re only talking about relative to other people in their own party here)

This explains a couple of otherwise mysterious things. How is Sanders on track to win in Iowa and New Hampshire when he is losing so badly nationally? Well, because Iowa and New Hampshire are two of the whitest states in the country. And how come I keep hearing people say “I’m sure Sanders will win, because even though the media and Big Business support Hillary, everybody I know supports Sanders”? Well, are those people white? Is their entire friend group white? Do they live in very strongly white areas? Then Sanders probably has much higher support among their friends and neighbors than he does nationally.

This might just be a transitory matter of two candidates with different styles and no relevance beyond this particular primary. Or it could represent the first cracks in the alliance that makes up the Democratic Party.

Racially, the Democrats are more diverse than the nation as a whole; since few nonwhites are Republicans, the Democrats are 60-40 white/minority. Socially, the Democrats combine enlightened college-educated creative professionals who want to help the poor, with poor people who want to be helped. And ideologically, the Democrats combine old-school quasi-socialists very concerned about Big Business and income inequality, with social justice activists who think the real issues are race and gender. So far these have been very benign splits. Everyone’s interests basically line up the same and nobody has a lot of reason to fight with anyone else – unlike the Republicans, who are already in civil war.

But the current election brings all three splits into near-alignment. The quasi-socialists, whites, and enlightened professionals generally support Sanders. The social justice activists, nonwhites, and poor people generally support Clinton – this bizarre situation of the guy most vocal about helping the least fortunate getting support from everyone except the least fortunate themselves. While I don’t really expect any fireworks to fly, it’s a risky situation and makes this an interesting time to be watching politics.

But mostly I bring this up not because the presidential primary is interesting in itself, but because it really drives home two important points that I’ve tried to make before.

First, in this post, I suggest that when talking politics “white” sometimes literally means people of European descent, but other times means what I dubbed the “Red Tribe”, very loosely corresponding to Republican voters, but also with connotations of southern, poor, uneducated, religious, and exaggeratedly patriotic. This seems to be one of those second times. Even if Donald Trump had 100% support from all minorities, he would still be “the white people candidate”, or even, as some people have called him, “the white power candidate”. Likewise, even if 100% of Sanders’ supporters were white and no black or Hispanic person had ever had the tiniest positive thought about him, we would never get the same kind of “is Bernie Sanders a demagogue harnessing white voters?” story that Trump inspires every day. Sanders supporters aren’t white! They have degrees from Ivy League colleges! They’re the good guys!

Second, in this post, I argue against the theory that groups with few black members are necessarily racist or exclusive (frequently seen as “Silicon Valley is problematic because of how few black techies there are”). I note that black people are severely underrepresented in groups as diverse as runners, BDSM participants, atheists, fanfiction readers, Unitarian Universalists, furries, and bird watchers. They’re also underrepresented in movements with apparently impeccable leftist and anti-racist credentials, like Occupy Wall Street and the US Communist Party. Given the frequency with which the “your group has few minorities, that means you’re racist and need to become more explicitly leftist in order to shrieve yourself” argument gets used to punch down at nonconformist or “weird” groups, there can never be too many counterexamples. And Bernie Sanders’ campaign is such a counterexample. It fits poorly with the “low nonwhite representation is caused by insufficiently strong social justice orientation” theory, but very well with the counter-theory I propose in that post: nonwhites are just generally less eager to join weird intellectual signaling-laden countercultural movements.

I take immense schadenfreude in imagining the people who like to write thinkpieces that “call out” polyamory or atheism for their insufficient minority representation, fidgeting and sweating and trying to justify their support for Sanders. I deeply enjoy the thought of them reading the article on ‘Berniebros’ (warning: possibly literally the worst article ever written, I am not kidding) and maybe realizing that wait, this is what they’ve been doing to other people all along, and it’s kind of unfair and hurtful. I mean, this will never happen. But it makes me happy to think about.

Bernie hasn’t done much specific to upset minorities; I doubt those stunts by the Black Lives Matter protesters mattered much one way or the other. And I would naively have expected his message of income equality and helping the least fortunate to go over better with people who are pretty unequal and unfortunate. And although Bill Clinton was pretty popular among nonwhites, I don’t see anything super-special about Hillary that would make her attractive to them.

So I think the explanation here might just be the same explanation as with the atheism and BDSM: Hillary has better name recognition and is more mainstream and less weird, in the same way not-BDSM and not-atheism are more mainstream and less weird. This is also my explanation for Trump’s relative success with minorities: he’s a household name in a way that Marco Rubio and Scott Walker aren’t, and he has vague good associations of strong leadership and economic savvy among the TV-viewing public.

And if Sanders supporters accept that in their own case, maybe they’ll be more understanding when other people plead the same.

# Contra Huemer On Morals

Michael Huemer thinks there are objective moral truths, because we’ve been moving in toward a particular coherent ethical perspective for the past few centuries, and for all we know this could be because that ethical perspective is Objective Truth.

Achitophel: That’s a pretty uncharitable way of putting it.

Berenice: But does this view really deserve more charity? Suppose I said that in the past, almost nobody wore ties. Now lots of people do. This is probably because ties are the objectively correct fashion choice.

Achitophel: What if people in a dozen different civilizations independently converged on wearing ties? Wouldn’t that provide much stronger evidence?

Berenice: People in a dozen different civilizations have converged on wearing ties. Go to France, Russia, China, or Nigeria, and chances are that the most important people you meet there will be wearing ties. Sure, the convergence isn’t independent, but neither was the convergence in values. You don’t think that India becoming a bicameral parliamentary democracy with a bill of rights had anything to do with Britain being a bicameral parliamentary democracy with a bill of rights?!

Achitophel: You’re trying to make it sound like imperial Britain forced their values down India’s throat. And maybe they did. But how come things like representative government, human rights, and decreased torture took off in a bunch of countries that were never colonized at all?

Berenice: Which countries?

Achitophel: Japan? Russia? China?

Berenice: Japan requires an overly restrictive definition of “never colonized”. And China and Russia require a frankly insane definition of “representative government, human rights, and decreased torture taking off”.

Achitophel: Not in an absolute sense! Relative to before!

Berenice: Give me the Yongle Emperor over Mao any day of the week.

Achitophel: Mao was bad. But he pretended not to be. He didn’t say “Let’s go kill a bunch of people because killing is glorious.” He said “We shouldn’t kill people, but sometimes we have to.” He didn’t say “You’re all my slaves, because I have divine right.” He said “We’re all going to work towards freedom together, but the best way to do that is by doing what I say.” He still had more liberal values than the Yongle Emperor, he just did evil despite them.

Berenice: I feel like this is an odd distinction to insist upon when you are sitting atop a pile of skulls.

Achitophel: And Xi is better than Mao.

Berenice: Not too different from Yongle, honestly.

Achitophel: All right. Fine. Let’s forget about independent development by different civilizations. Let’s say we’re mostly talking about the West – which remember, is still a lot of different countries. Britain. France. Germany. Italy –

Berenice: I am aware which countries are in the West.

Achitophel: These countries all converged on the same couple of values. And those values were all coherent with one another. It seems pretty clear that “emancipation of slaves”, “”freedom of speech”, “decolonization”…

Berenice: Wait a second. Sure, we’ve done a lot of decolonizing the past fifty years. But we did a lot of colonizing the five hundred years before that. In fact, around 1450 the West switched from barely colonizing at all, to colonizing lots of stuff all the time. If Huemer had lived in 1750, wouldn’t he have argued that the arc of the moral universe is long but it tends toward colonialism? And then declared colonialism an objectively correct moral truth?

Achitophel: Stop interrupting! “Emancipation”, “freedom of speech”, “decolonization”, “women’s rights”, and “democratic governance” are all kind of in the same moral direction, so to speak. Do you agree that Western values, today, not in 1750, TODAY, are all going in a certain coherent direction instead of varying randomly?

Berenice: You know, it’s not just ties.

Achitophel: What?

Berenice: If you think about it, practically every item of clothing has become less ornate. Think of Louis XIV in his huge expensive wig, his shiny blue fleur-de-lis filled fur robes, his carefully sculpted gold cane, his bejeweled ceremonial sword, his shiny red heels encrusted with diamonds, his gigantic outrageous hat, all sorts of weird neckbands and armbands. The Yongle Emperor would have had a more Chinese style, but it wouldn’t have been so different in conception. But nowadays nobody does that, not even the rich people who could afford it. The only time you’ll get shiny jewel-filled robes and fifty different things going around your neck is when somebody wants to look old-fashioned and traditional, like a Pope or Cardinal. And this is true everywhere. De Gaulle dressed more simply than Louis, and Mao dressed more simply than the Yongle Emperor. And when we picture the future, everyone’s dressed in featureless skin-tight suits. Evidence for objectively correct fashion?

Achitophel: There’s probably some driving force that made simplicity of clothing desirable, and which applied equally everywhere. For example, ornate clothing was a good signal of wealth back in Louis’ time. But after the Industrial Revolution, anyone could wear ornate clothing. Once the middle-class starts showing up to their bear-baitings in ornate fleur-de-lis gowns, wearing it just meant you were too clueless to know that it had no value anymore. So countersignaling took over – haven’t we talked about this before? The clothing thing isn’t because of some objectively correct fashion choice, it’s just a side effect of increasing wealth?

Berenice: Ding ding ding! Gold star for you! But why don’t you follow your theory to its logical conclusion and realize that the change in morality is also an effect of increasing wealth? Robin Hanson has just written about this in response to Huemer. Here, I’ll quote him for you:

One of the two main factors by which national values vary correlates strongly with average national wealth. At each point in time, richer nations have more of this factor, over time nations get more of it as they get richer, and when a nation has an unusual jump in wealth it gets an unusual jump in this factor. And this favor explains an awful lot of the value choices Huemer seeks to explain. All this even though people within a nation that have these values more are not richer on average.

The usual view in this field is that the direction of causation here is mostly from wealth to this value factor. This makes sense because this is the usual situation for variables that correlate with wealth. For example, if length of roads or number of TVs correlate with wealth, that is much more because wealth causes roads and TVs, and much less because roads and TV cause wealth. Since wealth is the main “power” factor of a society, this main factor tends to cause other small things more than they cause it.

This seems obviously correct to me and I don’t know why you and Huemer can’t see it.

Achitophel: You didn’t quote Huemer’s response! Here:

Perhaps there is a gene that inclines one toward illiberal beliefs if one’s society as a whole is primitive and poor, but inclines one toward liberal beliefs if one’s society is advanced and prosperous. Again, it is unclear why such a gene would be especially advantageous, as compared with a gene that causes one to be liberal in all conditions, or illiberal in all conditions. Even if such a gene would be advantageous, there has not been sufficient opportunity for it to be selected, since for almost all of the history of the species, human beings have lived in poor, primitive societies.

Berenice: Which gene that inclines us to take an airplane when we want to get somewhere quickly, but inclines us to take the bus if economy is more important? Is it DRD4 or SERT? I always forget that one.

Achitophel: You’re saying that it isn’t genetic.

Berenice: Or differently genetic, or complicatedly genetic, or gene-environmental-interactionic. This is what Robin Hanson says:

Well if you insist on explaining things in terms of genes, everything is “unclear”; we just don’t have good full explanations to take us all the way from genes to how values vary with cultural context. I’ve suggested that we industry folks are reverting to forager values in many ways with increasing wealth, because wealth cuts the fear that made foragers into farmers. But you don’t have to buy my story to find it plausible that humans are just built so that their values vary as their society gets rich.

Achitophel: That’s your argument? “We just don’t have good full explanations to take us all the way from genes to how values vary with cultural context?” Your whole point is just an argument from ignorance? Forgive me if I wait until you can come up with a plausible mechanism.

Berenice: You want plausible mechanisms? I’ve got your plausible mechanism RIGHT HERE. To put it in Haidtian terms, the Purity moral foundation, plus a sort of ethnocentrism that corresponds roughly to his Loyalty and Authority moral foundations, are carefully evolutionarily regulated by the prevalence of disease. Purity is the most obvious, given that the disgust reflex is obviously an evolutionary defense against pathogens. The reason you’re grossed out at the thought of touching feces, blood, or rats is that they’re full of plague; the reason you’re even more grossed out by the thought of eating them is that eating things is an even better way to get plague than touching things. Likewise, the best reason to avoid strangers is that they might have strange germs; about twenty million Native Americans who learned that lesson the hard way. Humans have an evolved behavior of upping their levels of purity and ethnocentrism under germ threat. Invent sanitation and antibiotics, eliminate most germs, and people naturally tend toward lower purity-concern and ethnocentrism. You get less racism, more sex, nontraditional families, cultural mixing, and all that good stuff. That’s why you get great correlations between the levels of pathogens in a region and the moderrness of their values. Go somewhere cold and lifeless like Sweden and you’ll get a liberal utopia. Go to a jungle in the Congo full of creepy-crawlies and everyone will be slashing everyone else with machetes. Really, read the article!

Achitophel: You think antebellum Southerners didn’t like black people because they thought they had cooties? Forgive me if the whole enslavement thing doesn’t seem to follow.

Berenice: I’m not saying that’s the only explanation or even the main explanation. You asked for a possible mechanism. I gave you one.

Achitophel: Fine. Give me a mechanism that explains slavery, then. And don’t you dare say it’s not the main explanation afterwards. Give me the best you’ve got.

Berenice: Have you ever noticed how much more virtuous rich people are than poor people? Poor people shoplift all the time, but rich people almost never do.

Achitophel: I don’t know where you’re going with this, but rich people commit white-collar crime and defraud people out of millions of dollars.

Berenice: Which just goes to show their moral superiority all the more! The poor person sells his principles for a dollar; the rich person holds fast until the temptation becomes absolutely overwhelming.

Achitophel: Shut up and make your point.

Berenice: A lot of moral decisions are a conflict between a principle and a temptation. People with fewer temptations have an easy time looking more principled. Not shoplifting is easy for a rich person, not because they’re more virtuous, but because they’re not in a position where they gain anything by doing so.

Achitophel: And this relates to slavery how?

Berenice: I would argue that we have many different drives and needs, some of which can be raw materials for making morality. Compassion is a drive. Xenophobia’s also a drive. Either one can be emphasized or deemphasized based on what’s useful or practical. If the most important thing for you is coming up with an excuse to enslave other people to make cotton, you might cultivate this primitive xenophobia into a complicated system of institutionalized racism that becomes the value system of your entire culture. If you’re not doing that, maybe compassion wins out. I mean, isn’t it interesting that all of the moral decent liberal people were north of a certain imaginary line, and all of the immoral bigoted people were south of it? And that imaginary line just happened to separate the climate where you could grow cotton from the one where you couldn’t? I’d argue instead that given a sufficiently lucrative deal with the Devil, the South took it. The Devil didn’t make the North an offer, and so they courageously refused to yield to this total absence of temptation.

Achitophel: You make the Southerners sound pretty Machiavellian.

Berenice: No more than the rest of us. I expect that once somebody invents vatburgers, we’ll all gain an sudden respect for animal rights, and recoil in horror that we ever engaged in factory farming. Until then, we come up with various moral justifications for the thing we’re not going to stop doing.

Achitophel: So liberal values are real morality, and older values are just excuses to justify greed?

Berenice: Not necessarily greed. “Necessity” is too strong, “convenience” is too weak, but somewhere in between the two. Back in the old days nobody really knew what STDs were. They just knew if you had sex too many times, you would break out in a horrible pox and die. And so would anyone else you had sex with, no matter how otherwise-pure they were themselves. Under those circumstances, having a very sex-negative morality where the promiscuous people are shunned and driven from society is a basic concession to the survival instinct. You’d be insane not to. But once we figured out testing and pencillin, the reasoning behind that morality died out and we stopped trying to cultivate those values. The sex-negative morality isn’t trying to justify greed. It’s making basic concessions necessary for survival. And you know what? If we suddenly had a zombie apocalypse and all of the gains of civilization evaporated, we’d be back to the old illiberal morality in the blink of an eye.

Achitophel: It still sounds kind of liberal modern values are the real morality, and other values are just sort of necessary evils.

Berenice: I think it’s more symmetrical than that. A lot of modern values would disappear if we stopped facing modern problems. We worry a lot about racial sensitivity, but if we ever got a society where racism was as thoroughly neutralized as syphilis, we’d probably drop that value pretty quickly too. If we ever totally conquer poverty, so that everyone’s got more than enough, maybe we’ll even stop worrying about compassion and fairness. Likewise, a lot of the democratic values – freedom of speech, freedom from slavery, equality, etc – are based on most countries being democracies which in turn is based on the historical situation. One of the big shifts was from the medieval system of “mostly super-well-trained professional warriors ie knights matter in projecting military force” to “any warm body with a gun matters”. That gave the common people a new level of power and probably led to democracy and the democratic virtues of equality and freedom. Likewise, technology has connected the world to the degree where different races and cultures and ideas are frantically mixing and mutating, making things like tolerance and freedom of thought much more relevant.

Achitophel: What about not torturing people? What about trying to solve poverty?

Berenice: So we’re too egalitarian to worry much about Authority and Loyalty. We’ve got too many antibiotics and contraceptives to care about Purity. But Care/Harm and Fairness seem as relevant as ever. Maybe even moreso. Given the advances in journalism, communication, and art, we have the ability to learn about and appreciate the struggles of others in a way we never have before.

Achitophel: That sounds a little forced. I could come up with a counter-story where given the worldwide increase in wealth and our lack of real-life exposure to any starving people or smallpox victims, the Care foundation atrophies away, but given our increasing crowding and exposure to superplagues like HIV and Ebola, Purity becomes obsessively important.

Berenice: \*shrug\* Maybe Care/Harm really is just the fundamental moral foundation, and the others are epiphenomena to be abandoned as we outgrow them. How does that saying go? – “The last enemy to be destroyed is submaximal global utility; destroying Death just buys us more time.”

Achitophel: So you kind of agree with Huemer after all?

Berenice: Perish the thought! Huemer thinks that this change in values proves there’s an objective morality and we’re moving toward it. The strongest claim I would dare is that one of these axes has always been the one that, all else being equal, would dominate the balance – and this is just the first time all else has been equal.

# Meat Your Doom

[Epistemic status: Very dirty and approximate, but I think roughly correct. Check my calculations and tell me if I’m wrong.]

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.”

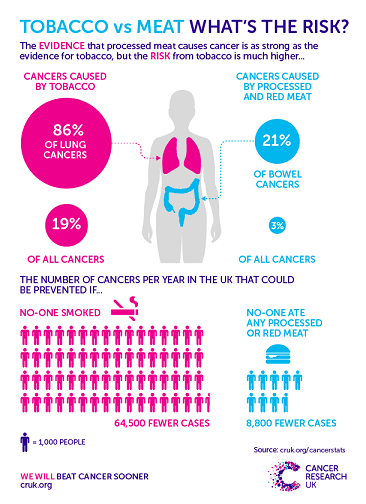
“But I don’t want to die!”

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:



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.

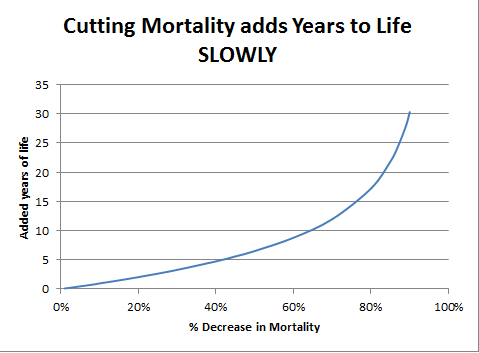
II.

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:



This is decrease and we’re talking increase, but 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)

# What Developmental Milestones Are You Missing?

[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 oeprational 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 an 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.

# Looking A Gift Horse In The Mouth

I started criticizing social justice back in 2010, which doesn’t sound so impressive until you realize that’s two centuries ago in Internet Years. At the time, you rarely heard such criticism outside of wingnutty lesbianism-causes-witchcraft circles. It felt bizarre, transgressive, and novel.

But over the past few years I’ve been privileged (sic) to meet many other people with the same concerns. Some were kindred spirits. A few at least had interesting ideas. Many others were horrible people next to whom the lesbian-causes-witchcraft types looked like Voltairesque voices of reason.

But they all had something in common: they were nobodies, and nobody cared what they thought. The lesbian-causes-witchcraft types had their talk shows, but among moderate liberals social justice criticism stayed mostly confined to a bunch of small blogs.

Now that’s over – during the past year big national media have unleashed a flood of social-justice-critical stories. The Atlantic published The Coddling Of The American Mind. Salon (Salon!) published Campus PC Panic Is Getting Ridiculous and How Coddled Young Radicals Got Discomfort All Wrong. The New Republic published Trigger Happy. Even President Obama has condemned what he called “coddled” college students, saying “that’s not the way we learn”. The UK political class is up in arms about Germaine Greer being denied platform, and the US political class is up in arms about the Halloween costume argument at Silliman College (nominative determinism!) in Yale. Complaining about social justice seems to be getting, dare I say, almost trendy.

As the old saying goes, “don’t look a gift horse in the mouth”. But as the other old saying goes, “I know some Trojans who would be a lot happier if they had.” So let me explain why this sudden outpouring of support for my position makes me uncomfortable.

When I or some other random blogger complains about the social justice movement, we tend to worry about points like the following (I won’t prove/defend these claims here, just clarify what I’m worried about):

Meanwhile, when important public figures and nationally circulating magazines complain about the social justice movement, I usually see language and arguments more like the following:

These seem like different agendas. In particular, the nobody-blogger angle focuses on ways in which social justice is used to justify aggression, and the mass-media angle focuses on ways in which social justice is used to coddle weakness. Thus the national magazines’ focus on trigger warnings, which happen to be one of the pieces of social justice I really like and have defended at length precisely because they do sometimes help weak people.

But there’s another common thread to the mass-media criticism: they’re all about things that happen on colleges and inconvenience college professors. Compare the recent bullying of a fan-artist to the point of suicide because she drew a cartoon character too thin and so was “erasing fat people” – to the recent students at Yale getting angry at an administrator who said she wasn’t going to enforce cultural sensitivity on Halloween costumes and so yelling and throwing stuff at her and her family.

I feel for anybody who gets yelled at and has stuff thrown at them, but the first of these two stories seems by far the most important; lots of teenagers commit suicide every year because of bullying, the idea that somebody deserves to die because they picture a cartoon character differently is abominable, and anyone who’s been on the relevant parts of the Internet knows this kind of thing is common as dirt. If I were a news editor, I’d consider the first study a much bigger deal. Instead, the second has gone viral in the national media, and the first remains stuck among the same few second-tier sites and SJ-critical nobody bloggers whom these kinds of things are always stuck among. Why?

I worry that the media, especially the online thinkpiece media, overrepresents an insular demographic of Ivy League academics and their friends who spend most of their time on college campuses and don’t notice things that don’t affect them personally. When people on Tumblr are being bullied to suicide or told that they’re garbage or outed or getting death threats, that’s the commoners. When a Contemporary Perspectives On American Literature professor is inconvenienced, AAAAAAAAH SOCIAL JUSTICE HAS GONE TOO FAR! SOMEBODY WARN SALON.COM!

Or to be even more cynical: social justice was supposed to be Yale’s weapon against Caltech and Podunk. But now Yale students are using it against Yale professors and administrators, and now it’s a problem. It’s like the police beating up city council members with the truncheons they usually reserve for poor ghetto-dwellers; you can bet there will be a newfound concern about police brutality at city council meetings.

And on the one hand, anything that inspires discussion of police brutality at city council meetings is good. Certainly the SJ-critical movement has been stuck on the same side as people a whole lot creepier than self-serving humanities professors, so what’s the problem?

I think that is the problem. When creepy white supremacists criticize social justice, they’re at no risk of taking over the wider SJ-critical movement. As the old saying goes, white supremacists are the best argument against white supremacy, and most of them couldn’t take over a blanket fort with a flamethrower. But rhetorically-gifted Yale professors who get thinkpieces published in The Atlantic are exactly the sort of people who would take over the wider SJ-critical movement, become its most important voice, and define what it means both to the rest of the world and to its own members.

That would be a disaster. Any general knows that you want to hold the high ground, and it really really shouldn’t be hard to hold the high ground against the sorts of people who continue to defend bullying someone to suicide because they drew a cartoon character differently than other people. But the mass media seems determined to find a way to yield the high ground and insist against all evidence that it’s punching down. Yale administrators might be the only group more sheltered than Yale students, yet they’re the only group the media seems to have the energy to defend. Worse, media seems to be defending them in a way that attacks activists for being weak and defenseless instead of pointing out when they’re strong and abusing their power.

I’m not saying that there aren’t important arguments to be had about trigger warnings and safe spaces. There are. But they’re only one of many problems, and far from the worst. And if people must focus on trigger warnings and safe spaces, I wish they would use one of about a zillion good arguments that don’t involve the pseudo-Nietzschean “You’re all babies! Stop crying, little babies!” tone. All this is doing is granting social justice activists their most dubious claim: that they are trying to use their ideology as a shield for themselves rather than a sword against others (as Popehat brilliantly puts it).

Finally, I think this might be a wake-up call to worry about the role of academia in media more generally. A friend on Tumblr pointed out that Hillary Clinton’s official list of campaign priorities include “ending sexual assault on campus”? Why not just “ending sexual assault”? Studies find that women are less likely to be assaulted on college campuses than off them. Isn’t “ending sexual assault on campus” the same kind of priority as “ending murder in gated communities?” Every murder is a tragedy, and murders in gated communities are no exception. But wouldn’t it reveal a lot about who mattered in a society if “end murder in gated communities” was how they framed their anti-murder initiatives?

I worry recent criticism of social justice is revealing the same thing.

# Hardball Questions For The Next Debate

Dr. Carson:

One of your most important achievements as a neurosurgeon was inventing the functional hemispherectomy, a treatment for epilepsy in which the epileptic hemisphere of the brain is severed from the healthy hemisphere and the body, allowing the healthy hemisphere to have full control of the body free from any epileptic interference. Children who get a functional hemispherectomy sufficiently early will be partly paralyzed on one side, but they will mostly be seizure-free.

Standard hemispherectomies remove the epileptic hemisphere from the body, but that tended to cause hydrocephalus, so your technique instead just severed all of its sensory and motor connections, leaving it present but inert.

But an anonymous neuroscientist on Reddit expressed some concern that just as the functional hemisphere seems to develop full independent personhood after the split, so the epileptic hemisphere may do so as well. Obviously it remains impaired by the epilepsy, but it’s not seizing all the time, so there will still be comparatively lucid intervals.

So my question for you is – what do you think happens to that person who is in an empty hemisphere, locked out of all sensory input and motor control? Do you think they’re conscious? Do you think they’re wondering what happened? Do you think they’re happy that the other half of them is living a happy normal life? Do they sit rapt in unconditioned contemplation of their own consciousness like an Aristotelian god? Or do they go mad with boredom, constantly desiring their own death but unable to effect it?

Ms. Fiorina:

One of the issues that’s played a central role in your campaign is your belief that the Ottoman Empire was the greatest civilization in the world. Certainly their five-hundred-plus year reign was marked by impressive military, political, and artistic achievements. But I want to bring up a particular aspect of Ottoman governance today.

One of the really unique Ottoman innovations was its so-called “millet system”, where every ethnicity and religion was almost its own little empire-within-an-empire. For example, although the Ottoman Empire was itself Muslim, Christians within it got their own millet, led by the Patriarch of Constantinople. They made their own laws, which applied only to Christians, settled disputes between two Christian claimants, levied taxes from Christians to pay for Christian-related projects, and generally kept their own people in line. When the Ottoman Empire as a whole wanted something from its Christian population, the Sultan would meet with the Patriarch and they would hammer it out. There were similar structures in place for Jews, Armenians, et cetera.

The past few years have seen an almost unprecedented rise in identity politics in America, usually marked by the claim that the society is using its weight to kick around people of some identity or another. Society is kicking around blacks. Society is kicking around conservative Christians. Society is kicking around bisexuals. They all feel like they’re getting the short end of the stick, but a lot of their preferences are mutually exclusive, and it’s hard to imagine some kind of centralized government policy that could satisfy any of them.

As an admirer of the Ottoman Empire, you’d be in a uniquely good position to import some of the advantages of the millet system into the modern Western world. Obviously this would be complicated given all the conflicting identity claims and the close quarters in which everyone is intermingled, but there are already some visions of what it could look like – including my own Archipelago – and if it were raised to the level of a national discussion, people could no doubt come up with many more.

So my question for you is – weren’t you a pretty crappy CEO?

Mr. Bush:

Assume that fitness-to-be-President is a normally distributed trait with known heritability. Suppose also that past elections have 100% efficiency; that is, they always choose the most qualified candidate. We can then use some of the standard regression-to-the-mean equations to determine the chances that the highest fitness-to-be-President individual in generation G will be the offspring of the highest fitness-to-be-President individual in generation G-1.

The single most fit-to-be president man in a population of 300 million would be about six standard deviations above the norm. If that man breeds with the single most fit-to-be-president woman, and if in keeping with findings for other complex traits heritability is about 60%, we would expect their offspring to be about 3.6 standard deviations above the mean in fitness-to-be-president. One in every 2500 or so people is 3.6 standard deviations above average, meaning there would be at least 120,000 equally good or better presidential candidates than they in the United States.

How high would the heritability of presidential fitness have to be before there was at least a 10% chance that the offspring of the two most presidential Americans was himself presidential material? My calculations suggest about 90%, which is very high compared to what we know about similar traits – but actually not entirely outside the realm of plausibility.

But if a maximally-presidential man breeds with a woman who is less than maximally presidential, the odds fall precipitiously. Suppose that a maximally-presidential man breeds with a woman who is merely in the 99th percentile for presidential ability. Now given a heritability of 60% there will be three million Americans more presidential than their average offspring. Even given a 100% heritability, there is only a 1/73 chance that their offspring will themselves be worthy of the presidency.

So my question for you is: do you think Barbara Bush is an unrecognized political super-genius, or are there probably hundreds of thousands of Americans who would make a better president than you would?

Senator Cruz:

You were on your college debate team, and you were good at it. Really good. You won the national championships and you were pretty widely believed to be the best debater in the country. Quite an achievement. But my worry is – which is more likely? That the best debater in the country would also be the best choice for President? Or that he would be really really really good at making us think that he would be?

Don’t respond yet. Before you answer that question – well, before you answer any question – we’ve got to think about this on the meta-level. There’s a classic problem in epistemology. Suppose that we have a superintelligence with near-infinite rhetorical brilliance. The superintelligence plays a game with interested humans. First, it takes the hundred or so most controversial topics, chooses two opposing positions on each, writes the positions down on pieces of paper, and then puts them in a jar. Then it chooses one position at random and tries to convince the human of that position. We observe that in a hundred such games, every human player has left 100% convinced of the position the superintelligence drew from the jar. Now it’s your turn to play the game. The superintelligence picks a position from the jar. It argues for the position. The argument is supremely convincing. After hearing it, you are more sure that the position is true than you have ever been of anything in your life; there’s so much evidence in favor that it is absolutely knock-down obvious. Should you believe the position?

The inside view tells you yes; upon evaluating the argument, you find is clearly true. The outside view tells you no; judging from the superintelligence’s past successes, it could have convinced you equally well of the opposite position. If you are smart, you will precommit to never changing your mind at all based on anything the superintelligence says. You will just shut it out of the community of entities capable of persuading you through argument.

Senator Cruz, you may not quite be at the superintelligence level, but given that you’ve been recognized as the most convincing person out of all three hundred million Americans, shouldn’t we institute similar precautions with you? Shouldn’t your supporters, even if they agree with everything you are saying, precommit to ignore you as a matter of principle?

Senator Rubio:

When you became Florida’s Speaker of the House, one of the other men on stage here tonight, Jeb Bush, presented you with a golden sword, which he said was the “Sword of Chang”. He told you that “Chang is somebody who believes in conservative principles, believes in entrepreneurial capitalism, believes in moral values that underpin a free society. Chang, this mystical warrior, has never let me down.” You looked pretty excited about it.

Now, some might say that this all came from a giant misunderstanding. Back in the late 1940s, Mao Zedong’s victorious Chinese communists forced Chiang Kai-shek’s defeated Chinese nationalists to retreat to the island of Taiwan. The United States kept the peace in the the Taiwan Strait, mostly to prevent Mao from invading and finishing the job, but a common refrain in 1950s conservativism went that we should “unleash Chiang”; that is, advise Chiang Kai-Shek to go back across the strait and reconquer China. George H. W. Bush served as envoy to China, had to listen to this sort of stuff, and got annoyed enough at the “unleash Chiang” rhetoric that he would quote it ironically at bizarre times, like his documented habit of threatening that his serve would “unleash Chiang” on his tennis opponents. It’s unclear how we got from George H. W. Bush’s constant threats to “unleash Chiang” on people, to his son’s belief that Chang was a mystical conservative warrior. Maybe it was a joke, either Bush Sr. pranking Jeb or Jeb pranking you.

In any case, you hung the sword in “a place of honor in your office”. From that point forward, Jeb’s fortunes declined. He left the Florida governorship, failed to get any further high positions, and then ran a very lackluster Presidential campaign. But from that same point your own fortunes decidedly rose. You started a law firm, were appointed a professor, got elected to the Senate, and are currently running a spectacular Presidential campaign with most pundits betting on your eventual victory after Trump and Carson lose their shine. The connection between the transfer of the sword and the sudden switch in both your fortunes is so striking that even the Huffington Post, not normally a source for magic-sword-related journalism, wrote about it: Jeb’s Last Hope – Reclaim the Sword of Chang.

But here we have a conundrum: if there was never a mythical Chinese warrior named Chang, by what magic does this sword grant worldly success to its possessor and ignomious ruin to any who lose it? There is a legend that fits almost exactly: the tale of the Holy Lance, aka the Spear of Destiny, aka several other portentious sounding names. According to the story, this relic from Christ’s crucifixion grants victory to all who own it and swift ruin to all who lose it. Charlemagne was reputedly the first to make use of its power; he was unstoppable while he wielded it but died moments after dropping it during battle. The same pattern repeated with Frederick Barbarossa, then a host of other military leaders, until finally it passed to the Austrian Habsburgs. They realized its power, locked it away, and ended up winning the greatest empire in European history. Supposedly Hitler was obsessed with it, so much so that his fascination with the object inspired the depiction of Nazi archaeologists in Raiders of the Lost Ark, and he took it for himself after the Anschluss. As the war wound down, the relic caught the special attention of General George Patton, who brought it back safely to Vienna afterwards. But ever since that time there have been various rumors that it was a fake, and that Nazi sympathizers took the real Lance in preparation for the time when the Reich would rise again.

The book Secrets of the Holy Lance describes one possible route by which the artifact might have been smuggled out of Europe:

Reporters John Buchanan and Stacey Michael cite recently declassified documents from the US National Archives that indicate that Prescott Bush “failed to divest himself of more than a dozen enemy national relationships that continued until as late as 1951. Bush conducted business following the end of World War II with moving assets into the Nazi refuges of Argentina, Panama, and Brazil.

So Prescott Bush was involved in moving Nazi “assets” from conquered Europe to South American refuges, presumably including the true Lance. Far be it from me to impugn his business ethics, but I don’t remember Nazi refugees in Argentina becoming an unstoppable force aided by a weapon of legendary mystical power. On the other hand, I do remember Prescott Bush being elected to the United States Senate just a few years later. Then his son and the presumed heir of his property was elected US President. Then his son was also elected US President. I need not add that according to the the laws of genetics, the chance of this happening by coincidence is hundreds-of-thousands to one even assuming implausibly high heritability of the fitness-to-be-president trait. Then his other son starts rocketing up through the ranks right up until the moment he gave you the sword of Chang, a sword named after a weird Bush family in-joke about a Chinese mystical warrior who doesn’t exist.

I think we can start to sketch out a plausible explanation here. Hitler didn’t want the Holy Lance falling into the hands of his enemies, so he replaced it with a fake and hired Nazi-artifact-smuggler Prescott Bush to transport the real one to safety in South America. Bush realized what he had, handed the South Americans a second fake, and kept the real one for himself, reforging it from a lance into a sword to cover his tracks – an action entirely in character for Prescott Bush, whose other relic-stealing adventures include the theft of Geronimo’s skull. He died unexpectedly without getting the chance to explain the significance of the artifact to his son George H. W. Bush. But since it seemed like a sentimentally important heirloom, George took care of his father’s weird golden sword anyway. When his sons asked him about it he didn’t have a real answer, so he just made his favorite in-joke about “unleashing Chiang”, and they believed him. Then eventually it passed to George W, later on to Jeb, and then Jeb thought it would be a funny present to give you to honor your election as Florida speaker.

Obviously the Lance is a significant strategic asset for America, and I imagine if you were President then its aura of victory would apply to the country as well, much as the Habsburgs’ possession of the lance enlarged Austria-Hungary. However, its powers are generally held to come from the Antichrist.

So my question for you is, do you think it’s ethical to use your magic sword to channel the power of the Antichrist if that would ensure America’s military success?

Mr. Trump:

You are famous both for your vast corporate empire and for your tendency to name the pieces of that corporate empire after yourself. By my count there are six buildings named “Trump Tower”, ten named some variation on “Trump Hotel”, a Trump Building, a Trump Palace, and a Trump Estate. You founded a financial services group called Trump Mortgage, a modeling agency called Trump Model Management, a bottled water brand called Trump Ice, and a magazine called Trump Magazine. You also started an airline called Trump Airlines, a TV company called Trump Productions, a book series called Trump Books, and your own radio talk show called Trumped!. There are also several Trump-themed games, like Donald Trump’s Real Estate Tycoon and Trump: The Game.

Mother Jones wrote a great article on this last one. Trump: The Game seems to be a tacky Monopoly clone. Players move around a board and bid on properties, and when one of them gets locked out of bidding for a property the other player gets to say “YOU’RE FIRED” the same way you do on your show. The only way to get back in to a property once you’ve been fired is to use the game’s most powerful card, which has a picture of your face on it and is called “The Donald”.

My question for you is: WHY DIDN’T YOU CALL IT THE TRUMP CARD?!?!!!!111111111asdfdf

# SSC New England Meetup And Presentations Schedule

1. Tonight (11/19) I’ll be speaking at Yale on the subject of “How To Ruin A Perfectly Good Randomized Controlled Trial” and other aspects of interpreting scientific evidence. Probably. I’m currently stuck in New Jersey due to airplane mishaps, but 90% confidence I’m going to be able to make it on time. Event is at LC101, Linsly-Chittenden Hall, 63 High St, New Haven, at 7 PM. You’re invited.

2. Monday (11/23) I’ll be speaking at Harvard at Sever Hall 203 at 5 PM on the same subject. You’re invited to this one too.

3. Sunday (11/22) we’ll be having a Boston meetup (thanks to the Boston rationalist and EA communities for setting this up) at MIT, room 5-134 at 6 PM. There is a Facebook event page here. You are invited to all of these things but you are extra definitely invited to this one. Every time I try to have a meetup I specify that everyone is invited even if you are only a lurker and even if you don’t understand everything you see on SSC and even if you are not very interesting and so on and so forth, and every time people still say “I wanted to go but I didn’t because I’m afraid I wouldn’t have been welcome there because I’m not enough like the typical SSC reader,” and so now I have to repeat in bold that you are definitely welcome.

I won’t be around much for the next week or two while I do this stuff and see my family for Thanksgiving, so apologies for not responding to emails/comments, etc.

# Reporter Degrees Of Freedom

I.

A sample of Thursday’s talk at Yale:



These are four headlines describing the same study, Milkie, Nomaguchi and Denny (2015). The study found that of twenty or so outcomes, only three of them – all measuring delinquent behavior among teenagers – show significant effect from time spent with parents (and this result remains after Bonferroni correction). So Vox has a great argument for their headline. The National Post has an okay argument for their headline even though it’s kind of cherry-picked. The Washington Post just sort of reads between the lines and figures that if it’s not quantity of time that helps kids, it must be quality. And FOX also reads between the lines and figures that if moms spending time with their kids has no effect, the argument from opportunity costs suggests mothers are spending too much time with their kids.

None of them are completely outright lying. And indeed, most of the articles eventually explain what I just said, halfway down the article, in one or two short sentences that most readers will skim over. But the rest of the article uses the study to support whatever the news source involved wants it to support, and so people will come up with four diametrically opposed conclusions from this one study depending on which source they read.

II.

Here’s a study that I wasn’t able to include in the presentation because it just came out recently. As per the Rice University press release: Overweight Men Just As Likely As Overweight Women To Face Discrimination.

The paper included two studies. In the first, men went into stores either with or without fat suits and try to do some things – ask if there were job openings, ask for a job application, ask an employee for help, try to buy some things, et cetera. Then they measured the men’s success across both conditions to see if they had more trouble when they appeared overweight.

In the second, subjects were asked to rate videos of an employee giving a marketing spiel for a new product; once again, the employee was either wearing or not wearing a fat suit. They measured the subjects’ ratings across conditions to see if they ranked the overweight employees lower.

The first study only included men, and so could not possibly have determined whether men were more or less likely than overweight women to face discrimination. The second study actually did have both male and female employees involved, and although it really wasn’t their main interest, the researchers did a post hoc evaluation to find the effect in each sex. In all three of the outcomes where discrimination was found, women faced more discrimination than men. They didn’t significance-test the comparison, but just from eyeballing it, it was probably significant.

So a paper in which one study does not compare men to women, and the other study finds women facing more discrimination than men, the press release somehow gets phrased as “Overweight men just as likely as overweight women to face discrimination in retail settings”. Huh.

You might wonder, “Does it really matter what a press release says? Does anyone read the exact wording?” Yes. Many other news sources copied the phrasing, for example Medical Daily’s Fat Discrimination Is The Same Regardless Of Gender. One such copycat, PsyPost.com copied the press release nearly word for word, including the title. Then it got posted on Reddit and now has 5189 upvotes and 1572 comments. So there’s that.

III.

“But at least it correctly raised awareness of how weight discrimination is a big problem in the retail setting, right?”

The paper measured a ton of different outcomes. Let’s focus on Study 1. The actor in the fat suit was supposed to ask if there were job openings (there were) and see if the company told him. Then he was supposed to ask for an application form and see if they gave it to him. Then he was supposed to walk in as a customer and see if employees greeted him. Then he was supposed to ask the employees to recommend him an item and see if they did. Then he was supposed to ask them to recommend him a second item and see if they did.

No difference was found between overweight and normal-weight actors in any of those five experiments. Two of them had ceiling effects that probably made the attempt futile, but the other three didn’t, and there wasn’t even a trend toward discriminating against the overweight guy.

So what did they find discrimination on? They say that detected “interpersonal discrimination”, ie discrimination based not on any quantifiable outcome but based on how friendly/warm the person interacting with the actor seemed toward him. They determined this by self-rating and other-rating; that is, the actor wrote down how friendly he thought the store clerks were toward him, and a spy surreptitious observer who had placed herself near the interaction also rated this for corroboration. Their rating scales included twelve items including “how many times did the clerk nod”, “how friendly did the clerk seem?”, “how much eye contact was the clerk making?” and “how much comfort level did the clerk seem to have?”. The experiment found a statistically significant difference between the fat-suit-wearing and non-fat-suit-wearing trials and concluded that there was interpersonal discrimination.

But hold on a second! The study says nothing about anyone being blinded. In fact, it’s really hard to blind an actor to the fact that he is going into some stores while wearing a fat suit and other stores while not wearing a fat suit. As far as I can tell, everybody involved was in on the study from the beginning. If your boss tells you “I want you to rate how much comfort level clerks have with you for this study on fat discrimination”, it seems really possible to me that there might be a slight tendency to overrate the clerks who interacted with thin-you, and to underrate the clerks who interacted with fat-you.

How slight a tendency? Clerks dealing with fat people got an average rating of 2.3 (seven point scale, lower is better), and those dealing with thin people of 2.0.

So after finding no discrimination on five objectively measurable outcomes, they find very subtle discrimination on an unblinded subjective outcome practically designed to produce placebo effects.

We move on to the second study, where participants (as usual, psychology students) are rating video presentations given by fat vs. thin people. This is supposedly tying into the “retail industry” theme of the paper, but honestly it seems kind of forced to me.

Anyway, the participants are asked to rate their presenters on seven measures: overall quality of presentation, overall attitude toward product being presented, overall attitude toward the store that would employ a person such as this, intention to support the store, employee’s appearance, employee’s carelessness, and employee’s professionalism. The results:

There was no difference between how participants rated overweight vs. normal-weight presentations overall.

There was no difference between how participants rated products presented by overweight vs. normal-weight people.

There was no difference between how participants rated stores staffed by overweight vs. normal-weight people.

There was no difference between how likely participants were to support stores staffed by overweight vs. normal-weight people.

There was a difference in how participants ranked the appearance, carelessness, and professionalism of overweight vs. normal-weight people.

The first four results are encouraging. What about the last three?

Well, I feel like if you ask people to rank someone based on “their appearance”, and your subjects answer based on how they look, you kind of walked into that one. Oh no, people rank conventionally attractive people as having better appearances than less conventionally attractive people! Someone call John Ioannidis to double-check this astonishing result!

“Carelessness” and “professionalism” are perhaps less excusable, but c’mon, you had them watch a two-minute video. When you give someone zero information on a thing, and you force them to make a judgment on the thing, then yes, stereotypes are their best source of information. If you showed me a picture of an average-looking man and an average-looking woman and say “Quick! Which of these people is more likely to like baking cupcakes?!” I’ll pick the woman, not because I think all women are obsessed with cupcakes or because I go around looking at every woman I see as a cupcake factory, but because you asked me a stupid question and ensured stereotypes were the only thing I had to go on.

Then the authors find that this was mediated by explicitly-expressed stereotypes against fat people, which is kind of interesting, but doesn’t make the nonsignificant things any more significant.

So to sum up: there was no discrimination against the overweight on any objective measure of the actual retail experience, including positions advertised, applications given, greetings offered, or customers served. There was also no discrimination against the overweight on presentation evaluations in terms of overall evaluation, evaluation of employee, evaluation of product, or evaluation of company.

There was a tiny amount of discrimination on a subjective measure rated by unblinded observers aware of the purpose of the study. There was also some evidence on three subtler ratings of the presentation that seemed designed to ask participants impossible questions in order to force them to stereotype. However, these meaningless scales did not effect the raters’ overall impressions as measured any of four different ways.

IV.

Here is the reporting from the news outlets that passed their first test and didn’t frame it as men and women facing equal amounts of weight discrimination.

Business Insider: Researchers Had Men Pretend To Be Obese – And The Results Are Disturbing, which says that “this research highlights the importance of including men in discussions about weight stigmatization,” and “the authors also advocate organizational efforts to combat negativity against heavy customers and potential employees…the first step may be for individuals to become aware of how strong weight biases are.”

AskMen: Young Men Who Appear Overweight Suffer Interpersonal Discrimination. “Researchers disguised six thin young men as obese customers or job applications and found that they were victims of microaggressions. Basically people were a little bit more jerky towards them.”

Oximity: Overweight Men Often Snubbed At The Mall. “Shopping malls can be hostile places for overweight men, regardless of whether they’re customers or simply looking for a job.”

The Health Site: Men, Here’s One More Reason For You To Lose Weight. “Ruggs said that these findings were another reminder that there was still more work to be done in terms of creating equitable workplaces for all employees, potential employees and consumers. She concluded that this was something organisations could take an active role in, and said that companies could do better job training on customer relations as part of the employees’ new-hire process. ”

Now I’m almost missing the kind of random scattershot media bias we found on the time-spent-with-children study. Here every media outlet reports the results the same way that the study’s author and the press release reports the results.

This is not a totally wrong interpretation, any more than “six hours a week will tame your teen” is a totally wrong interpretation of the childhood study. But if I myself were writing an article on this study, it would be SURPRISINGLY LITTLE DISCRIMINATION FOUND AGAINST OVERWEIGHT MEN, and mention somewhere in the middle that some discrimination was found on a few sketchy variables. Instead, we get DAILY DISCRIMINATION AGAINST OVERWEIGHT MEN and WE NEED TO INSTITUTE SENSITIVITY TRAINING PROGRAMS IN RETAIL ENVIRONMENT, and they mention somewhere in the middle that a lot of important variables came out negative.

A lot of studies work like this. You test ten or twenty complicated variables, you get positive results on some, negative results on others, some of those results seem plausible, other results seem like maybe you made a mistake somewhere or didn’t have enough power or whatever, and then you make an interpretation based on your personal bias. Then it goes from the researcher’s personal bias to the abstract to the press release to the headlines to the mind of the average reader, dropping subtlety at each step, until “No discrimination against overweight men, except where the study was practically designed to ensure false positives” becomes “Rampant discrimination against overweight men everywhere” becomes “Overweight men are discriminated against just as much as overweight women.”

Don’t get me wrong. I expect there probably is lots of discrimination against overweight men. And I think this study’s project of trying to find it and convince people of its existence was worthwhile. But I don’t think you should get to convince everyone that science has proven X, unless science has actually proven X. The process that produced these headlines is strong enough to produce any headline you want, with the part where you actually do the study becoming more and more of a ritual or a formality. There are just too many degrees of freedom between the study and the reporting.

Stalin once said that “those who vote decide nothing; those who count the votes decide everything.” It’s starting to look like those who do the studies decide nothing and those who report the studies decide everything. The only solution is to actually read the study and not just the headlines. Sometimes we might even have to – God help us – read beyond the abstract.

# College And Critical Thinking

[epistemic status: My bias is against the current college system doing much good. I have tried not to be bogged down by this bias, but take it into account when reading my interpretations below.]

[EDIT: An earlier version of this post claimed that one paper had shown a u-shaped relationship between time spent in college and critical thinking. A commenter pointed out this was true only of a subset in two-year colleges, but not of four-year colleges or college in general – which shows the expected linear relationship. I am sorry for the error, and correcting it somewhat increases my confidence in college building critical thinking.]

Over Thanksgiving, I was discussing tulip subsidies with the pro-Bernie-Sanders faction of my family, and my uncle claimed that we needed college because “it teaches you how to think critically”.

The evidence sort of supports him, but with the usual caveats and uncertainties.

First of all, what the heck is critical thinking? Luckily, we have a very objective scientific answer: critical thinking is the ability to score highly on the Watson-Glaser Critical Thinking Appraisal. Now that we’ve answered that, we can move on to the question at hand.

Most studies on this issue are terrible because they lack control groups. That is, they measure students when they enter college, measure them again when they leave college, and find that their critical thinking ability has improved. But this could be for any number of reasons. Maybe older people generally have better critical thinking than younger people. Maybe life experience builds critical thinking. Maybe college had nothing to do with any of it. The best meta-analysis of such studies, MacMillan 1987, finds exactly this, and concludes:

Overall these studies suggest that seniors, in the main, are probably better at critical thinking than freshmen. However, since the most compelling data were gathered through weak pretest-posttest or longitudinal designs, it is difficult to separate out the effect of college from the maturational effects that occur despite college.

I like the phrase “maturation that occurs despite college”, although I don’t think they meant it in that way.

But in any case we need a better study design to conclude anything from this. There are two studies with moderately good designs, both by a guy named Pascarella. The first compares 30 college students to 17 matched non-college students and follows them up for one year. They find that while both students and non-students gain critical thinking skills over the course of the year, the college students gain 17% more, corresponding to an effect size of 0.44.

The second, larger study compares students doing college full-time to students doing college part-time, under the theory that if college is causing the effect, then a little college should cause a small effect, but lots of college should cause a big effect. They find this in the four-year college sample, and a garbled u-shaped mess in the two-year college sample. At least the four-year sample, which is what most people are interested in, looks good.

On the other hand, some other studies find less impressive effect sizes. Arum and Roska recently wrote a book on this kind of thing, Academically Adrift, and they find that two years of college (start of freshman to end of sophomore) only increases critical thinking by 0.18 SD. This is weird because it’s twice as much college as the past few studies but less than half the effect size. Also, it doesn’t seem to be controlled, so this is the sum of actual-college-effect-size and confounder-effect-size. According to one review:

College entrance to end of sophomore (ie half of college) improves critical thinking by 0.18 SD. This might have been greater in the past “Pascarella and Terenzini estimated that seniors had an 0.5 SD advantage over freshmen in the 1990s. In contrast, during the 1980s students developed their skills at twice the rate: seniors had an advantage over freshmen of one standard deviation.”

Note that we’re comparing unlike with unlike. Four years of college need not produce an effect twice as great as two years of college, any more than a space heater that increases the temperature of a room 10 degrees after being left on for one hour will increase the temperature 87600 degrees after being left on for a year. Indeed, some studies suggest that most of the gains happen in freshman year. But even so, there’s a very clear downward trend here. As usual, we have no good way of knowing if that’s caused by gradually-improving studies, gradually-improving college critical thinking training, or gradually-deteriorating colleges.

One more question: do we know the specific aspects of the college experience that cause critical-thinking gains?

Specific explicitly-advertised “critical thinking classes” don’t. Being a college that prides itself on a specific “critical thinking focus” doesn’t. This sort of thing seems very well replicated, although there are a few individual studies that don’t really interact with the rest of the literature that seem to have at least temporary positive results. Classes that you might think would teach critical thinking, like logic, or philosophy, or statistics, don’t seem to do especially well (here’s a specific in-depth discussion of philosophy). The only positive result anyone’s been able to find is that “liberal arts” (here viewed as a broad category including science and mathematics) seems to do better than occupational skills, and “education, social work, communications, and business” seem to do worse.

In terms of broader factors, – one study finds that quantifiable college experiences explain “between 7 and 17%” of the variability in first-year critical thinking gains (other sources say classes explain 2.5% and extracurriculars 2.9%). Studying a lot seems to help. So does reading unassigned books. Aside from that, the biggest finding is kind of concerning:

Students who characterized their relationships with other students as “competitive, uninvolved…alienated” were more likely to show gains in critical thinking than were students who portrayed their peer relations as “friendly, supportive, or a sense of belonging” The data in this study do not permit confident explanation of this relation, but one might speculate that a sense of participation in a friendly, supportive peer environment may require a partial suspension of one’s critical thinking skills.

…wow. I was going to say something like “students busy spending time with their friends have less time to learn stuff”, but your cynical awful explanation works too.

So what’s the big picture?

Well, we know that people will gain critical thinking skills during the four years from age 18 to age 22. We have an small study that finds college helps a little with this process and a larger study that shows dose-dependent effects of college. We have some hard-to-compare effect sizes ranging from 0.18/2-years to 0.44/year. That’s modest but appreciable, and it’s probably at least somewhat real.

But those of you who went to my talk last week hopefully know what my next question will be: how long does this last?

For example, we know that parents’ personalities have all sorts of interesting effects on their children while those children are living with their parents. We also know that as soon as children leave their parents, those effects go down to near zero. What people tried to interpret as some deep fact about development was actually just a reflection of the environment that those children were in. Likewise, preschool makes children do much better in kindergarten, but by third grade the preschool-educated kids are doing the same or worse as the others. It’s not fundamentally altering their developmental trajectory, it’s just creating an short-term effect.

Every one of the studies I’m citing here was done in freshman or sophomore year (one study cited another done in senior year, but I couldn’t find it directly). No one has ever looked at students who have been out of college a year – let alone out of college thirty years – to see if the effect continues. I would bet that it doesn’t.

Until somebody checks, enjoy your opportunity to tell people that the evidence backs college building critical thinking skills.

# Setting The Default

[Epistemic status: I predict everyone except me will respond to this with “Duuuuuuuuh”, but I found it changed some views of mine]

I.

I recently did couples therapy with two gay men who’d gotten married a year or so ago. Since then one of them, let’s call him Adam, decided he was bored with his sex life and went to a club where they did some things I will not describe here. His husband, let’s call him Steve, was upset by what he considered infidelity, and they had a big fight. Both of them wanted to stay together for the sake of the kids (did I mention they adopted some kids?) but this club thing was a pretty big deal, so they decided to seek professional help.

Adam made the following proposal: he knew Steve was not very kinky, so Adam would go do his kinky stuff at the club, with Steve’s knowledge and consent. That way everyone could get what they wanted. Sure, it would involve having sex with other people, but it didn’t mean anything, and it was selfish for a spouse to assert some kind of right to “control” the other spouse anyway.

Steve made the following counterproposal: no. He liked monogamy and fidelity and it would make him really jealous and angry to think of Adam going out and having sex with other people, even in a meaningless way. He argued that if Adam didn’t like monogamy, maybe he shouldn’t have proposed entering into a form of life that has been pretty much defined by its insistence on monogamy for the past several thousand years and then sworn adherence to that form of life in front of everyone they knew. If Adam hadn’t liked monogamy, he had ample opportunity to avoid it before he had bound his life together with Steve’s. Now he was stuck.

Adam gave the following counterargument: yeah, marriage usually implies remaining monogamous, but that was all legal boilerplate. He had wanted to get married to symbolize his committment to Steve – committment that he still had! – and he hadn’t realized he was interested in fetish stuff at the time or else he would have brought it up.

Steve gave the following countercounterargument: okay, this is all very sad, but now we are stuck in this position, and clearly only one of the two people could get their preference satisfied, and given the whole marriage-implies-monogamy thing, it seemed pretty clear that that person should be him.

So then of course they both turned to me for advice.

The rules for psychotherapy are a lot like the rules for Aaron Burr: talk less, smile more, don’t let them know what you’re against or what you’re for. This last principle is generally known as “therapeutic neutrality”, and it demands that we not take sides in our patients’ disputes or dilemmas. Instead, we try to remain an impartial discussion facilitator, teasing out our patients’ true values and concerns until they are able to come to a conclusion on their own.

On the other hand, their friends probably don’t have the same scruples and are going to be offering them advice. And I wondered what advice I would give, if I were their friends and not stuck in metaphorical Switzerland.

II.

Assume Steve’s analysis is right; this is a zero-sum game and there’s no way for them both to come out of it happy. Which side do we choose?

A quick retreat to a simpler situation: suppose Adam really wants to keep all the windows in the house open all winter with no heat on, so that the inside temperature is 10F and the house is full of snow. Steve does not want to do this. Both of them want to stay together for the sake of the kids, but this do-we-freeze-our-house thing is really getting in the way.

This problem is easy. Adam, you’re crazy and your preferences are stupid and don’t count. Suck it up and keep living with Steve at normal-person temperatures.

Another retreat in the other direction: suppose Adam wants to sometimes take a shower, but for some reason the thought of Adam being in a shower pisses Steve off and he refuses to allow it. Once again, both of them want to stay together for the sake of the kids, but this can-Adam-take-a-shower thing is really getting in the way.

This problem is also easy. Steve, this time your preferences are stupid and don’t count. Suck it up and let Adam take a shower.

A third new situation. The one Unit of Caring recently discussed on her blog. A transwoman wants to have Christmas with her family, but her family doesn’t believe in transgender and insists on calling her by her original male name and male pronouns. Both her and her family don’t want to “ruin Christmas” by refusing to get together as a family or making a big deal of this. Who is in the right? Unit of Caring writes:

Anyone who does not respect their siblings enough to call them by the name of their choosing does not then get to go “oh! you not wanting to let me repeatedly hurt you is breaking apart our family, how unreasonable of you!” If you want people to spend time around you, call them by the names they chose. If you wouldn’t repeatedly slap your siblings in the face, don’t deliberately misgender them either. (And if you would repeatedly slap your siblings in the face, then you shouldn’t have to look too far to figure out whose behavior ruined Christmas.)

If it doesn’t bother you that you’re hurting someone, then you don’t get to act wronged when they decide you’re not worth spending time with.

I agree with this assessment, but only because I agree with Unit about the object-level issue of transgender. It seems like if you wrote in the same question to your local priest, they’d say the trans woman was being unreasonable. I don’t think there’s any good way for Unit and the priest (or the woman and her family) to resolve their differences except by one convincing the other of their position on the object-level issue of transgender.

(well, if you really really really understood utilitarianism, you might be able to say you should take the highest-utility solution, but no one understands utilitarianism that well)

This seems to be true of my patients’ problem too. Unless we can decide whether wanting to go to a fetish club and have sex with people besides your husband is a reasonable request, we can’t solve Adam and Steve’s disagreement. I mean, Steve’s argument about the contract isn’t bad, but if it were something we disagreed with – let’s say some old-timey marriage contract where the woman vowed to always serve and obey her husband, and now she’s a feminist and wants out – we would probably be pretty sympathetic despite the precise wording of what she’d “agreed” on.

III.

I come to the table with personal baggage. I come from a very permissive subculture. I’ve had some very happy open relationships and wanting to be open seems like a reasonable request. I’ve had some friends who are very kinky, and wanting to be kinky seems like a reasonable request too. I’m not personally very good at feeling jealous, so wanting your husband to never go to a club, even if he doesn’t tell you about it, or make you think about it, or even agrees only to do it when you’re away on a business trip in another city – seems a bit odd. Honestly I would be tempted to take Steve aside and ask him whether he’s sure that he couldn’t deal with Adam going to this club, and whether maybe he wants to give it a chance, and whether maybe he just wants what’s best for Adam even if that makes him a little uncomfortable.

But go back two hundred years and ask the people of that culture, and this choice is a no-brainer. Fetish clubs (or the closest 19th century equivalent) are weird, vile, sinful things, and Adam’s desire to go to one is totally beyond the pale. He should never even have made the request. But since he did, we can strongly and clearly tell him that this is morally wrong, that he should apologize to Steve for the trouble he put him in, that he should realize there’s more to life than kinky sex, and that he should want what’s best for Steve even if that means he can’t satisfy his libido quite so much.

If Adam and Steve were in the traditional culture of the 1800s there would be no debate. If they were in some ultra-permissive sexually-open subculture of the 2100s, there would also be no debate. The culture would tell one of them that they were wrong, just like someone who wants to make the other live in a 10 degree frozen house is wrong, that person would grudgingly agree, they would stay together, and that would be that. The problem only comes when they’re in a culture with a lot of different subcultures that haven’t made up their minds yet. Like ours.

We all hear the stories of the economists who start by assuming perfect rationality, and then add in deviations from that assumption when they come to them. I kind of like to start from a liberal assumption of perfect atomic individualism and add in deviations when I encounter them. And, well, this is the latest one I encountered.

Adam and Steve’s individual personalities and situations will help resolve their conflict, but the tiebreaker vote is always going to be cast by the culture around them. Realizing this has made me more open to activists who are trying to change the culture – and, symmetrically, to conservatives who are trying to prevent the culture from being changed. People with unusual sex lives like to say that what they do in the privacy of their own bedroom doesn’t hurt anyone else – neither breaks their nose nor picks their pocket – but the fact is that the partial social acceptance of fetish clubs and of open relationships is what gives Adam a leg to stand on. And some religious conservatives like to talk about how they only want to defend their own right to practice and express their beliefs instead of being forced into the broader cultural revolution all around them, but the fact is that their beliefs are what’s supporting Steve. My sympathies will always be with the atomic individualists who want to come up with some clever Adam-Steve contract that solves their problem on the meta-level as long as all actors are rational, but I am starting to worry the culture warriors have a point here.

UR said that “the sovereign is the one who sets the null hypothesis”. Once you’ve let the culture set a default – going to fetish clubs is a reasonable request, going to fetish clubs is an unreasonable request – then given sufficiently good liberal norms people who want to deviate from the default can absolutely do so, but as soon as a conflict springs up the identity of the default option still matters a lot.

I’m not suggesting a total war of all against all, and there’s always the Archipelago option, but I guess sometimes culture wars do need to be fought beyond the point where you just leave people alone, if only to shift the default in your direction.

Speaking of culture wars, an apology to gay people. I always obfuscate details about my patients to disguise their identities, but I feel particularly bad about making this couple gay because it reinforces the stereotype of gay people as hypersexual and bad at committment. I made them gay anyway, because when I tried to write them hetero, their gender seemed to skew the problem too much to one side or another – for example, when Steve was a woman, he was the poor innocent wife wronged by a horny husband who insisted on thinking with his crotch. I worried that if I made the couple hetero, my readers for one reason or another would bring their own baggage and wouldn’t be able to see it as the difficult and evenly-balanced problem it seemed like when I was in the office with them.

Which itself says something about how our culture sets default hypotheses.

# Product Recommendations 2015

It’s Christmas shopping season, and so time for the annual reminder that if you want to support this blog you can shop through my Amazon affiliate link (also on sidebar) and I’ll get ~5% of whatever you spend at no extra cost to you.

If that doesn’t appeal to you, you can also shop through this portal and give ~5% of whatever you spend to one of GiveWell’s top recommended charities.

In order to get you started, here are some recommendations for products I really liked over the past year. I’m not affiliated with any of these and don’t get anything from endorsing them besides the Amazon fee. I talk about some health care products, but they’re all out of my field and best thought of a the opinion of an informed consumer, not as official medical endorsements.

Traditional 18-Year Balsamic Vinegar

A big part of people’s enjoyment of food is placebo. For example, people in blind taste tests prefer Pepsi to Coke, but people in unblinded taste tests prefer Coke to Pepsi because of the Coke “mystique”. Likewise, people will rate wine as better-tasting if they are falsely told it is more expensive. I expect this effect is bigger in people (like me) with relatively dull palates and overactive imaginations. We could capitalize on this by finding some food that sounds magical, amazing, and precious, and expecting it to have a taste that matches.

That food is balsamic vinegar. Not the fake stuff you get for $4.99 at the supermarket. The real version, which is usually called something like “traditional 18-year balsamic vinegar” and whose manufacture involves about the same number of complicated restrictions and obscure types of wood as building the Ark of the Covenant.

It’s not just that the grapes have to be from a particular species and grown in a particular Italian soil. It’s that they then have to be aged through a process in which the vinegar is constantly transferred, untransferred, and retransferred among series of successively smaller barrels of oak, chestnut, cherry, juniper, mulberry, and acacia a bunch of times for eighteen years, so that each wood can “add its flavor” to precisely the right degree. How long it must spend in each cask is so complicated that people need high-level mathematics just to figure out how old any specific sample of vinegar is. See for example Guidici & Rinaldi 2007, A Theoretical Model To Predict The Age Of Traditional Balsamic Vinegar. The paper finally concludes that “there is a finite limit for the age of balsamic vinegar” – which I guess is reassuring.

So how does it actually taste? Kind of like a mixture of grapes, oak, chestnut, cherry, juniper, mulberry, and acacia It tastes really interesting. Definitely not like normal vinegar. Not really like normal anything. Italians put it on ice cream (really!) or drink it straight from the bottle (really!) I can’t promise it’s not just the placebo effect, and honestly it probably is, but it’s certainly a more interesting the placebo effect than just buying an expensive wine or eating at a fancy French restaurant.

Buy traditional 18-year balsamic vinegar

Fake Nice Pants

I have some sensory issues which make me find normal pants annoying; I prefer to wear sweatpants whenever I can. But this isn’t always compatible with appearing as a normal productive member of society, so sometimes I am forced to wear uncomfortable blue jeans or dress pants or whatever.

Luckily now there are are now sweatpants that look like jeans, khakis, dress pants, etc. I won’t say they’ll fool a careful examination, but hopefully nobody is carefully examining your pants, plus even if people notice they might feel awkward calling you out on it.

Buy sweatpants that look like jeans, sweatpants that look like other types of jeans, sweatpants that look like dress pants, or sweatpants that look like khakis.

Xylitol Nasal Spray

Xylitol nasal spray is for nasal congestion and allergies. It’s not too different from saline sprays, but it’s a little gentler on the body and also has mild antibacterial properties. Also, I hadn’t even realized how well saline sprays worked until recently.

I’ve informally recommended this to a couple of patients with nasal and sinus issues and have received mostly good reports from them too.

Buy xylitol nasal spray

Hamilton CD

The new Broadway musical Hamilton is really good. If you don’t believe me, believe mainstream media sources like 538, Vox, Breitbart, Vox again, the New York Times, Vox a third time, the LA Times, and one more Vox. Also all of Tumblr.

It’s a musical about the life of Alexander Hamilton set to rap, but you will like it even if you do not like rap or early American history. You can listen to the whole soundtrack for free on YouTube, follow along with the heavily annotated lyrics on genius.com, and finally you’ll want to have the CD so you can force your friends and family to listen.

If nothing else, this will help you understand the in-jokes on social media.

Buy the Hamilton CD

Dental Floss Scythes

Like everyone else in the world, my dentist told me to floss more but I never listened. Dealing with that weird little box of string was just too much of a trivial inconvenience.

I recently learned about dental floss scythes (probably they have a less interesting official name). These are little plastic things that kind of pre-arrange the dental floss for you. Now I am actually flossing every day. Well, most days. Some days. The point is, I’m flossing. Yes, I’m a bad person contributing to disposable waste culture, but at least I’m healthy.

(Now some people say flossing might not prevent heart attacks as was previously believed, but it’s probably still a good idea)

Buy dental floss scythes

Nootropics Depot and Ceretropic

Not a product so much as a company. Nootropics Depot and Ceretropic are two online nootropics stores. They’re both owned by the same guy but have different branding: Nootropics Depot has nice gentle all-natural stuff, Ceretropic has new high-tech research chemicals.

There are lots of online nootropics stores, but these are far and away the best. I say this after lurking on r/nootropics for a long time, where users are very vocal about their experiences with different companies and mostly agree with my assessment. The guy who runs these stores also hangs out on r/nootropics, where he shares his encyclopaedic knowledge of everything and helps people who have gotten themselves in trouble. There are worse ways to spend a day than just reading through his entire posting history, but his commentary on my article Iron Curtain Of Psychopharmacology is a good specific example. He is legit and so are his sites.

Experimenting with nootropics is fundamentally a risky endeavor. You are using experimental psychoactive chemicals that haven’t been rigorously tested for safety or efficacy. Nootropics Depot and Ceretropic don’t change any of that. But they do ensure you are not doing anything stupider than the stupid thing you think you are doing. They are very careful about having purity-tested, un-degraded chemicals and making sure you are getting what is advertised on the bottle. They try to have the gold standard version of everything – for example, if they’re selling an herbal extract, it will be the same by composition, subspecies, etc as the version used in whichever studies have most clearly suggested safety and efficacy for the herb. They avoid at least some of the sketchiest diet-pill-stimulant-amphetamine-analogs you can find on some other sites. And they have excellent customer service and if for some reason your nootropics don’t arrive or get confiscated by customs they will work it out for you.

The other reason I like this company is that it does seem to have a long-term plan to build up a power base from which to try to ‘disrupt’ the pharmaceutical industry in the way we often talk about here. That’s obviously crazy ambitious, but they’ve already done some impressive things, like come up with a longer-lasting version of tianeptine. From a chemistry point of view that may not be super hard, but since tianeptine is a prescription drug in most of Europe it means they’ve improved pretty substantially upon a real pharmaceutical company’s project. They seem to be gradually getting labs and researchers, so I’m optimistic.

Oh, and if some of you crypto geniuses want to do a mitzvah, find an easy way for them to charge people non-Bitcoin money without credit card companies backstabbing them every few months because they get cold feet and decide nootropics are too weird.

Visit Nootropics Depot and Ceretropic.

Periscope Glasses

In case the nootropics don’t work well enough for you and you still have human experiences like “sleep” and “tiredness”, you may be interested in periscope glasses. They let you lie down supine in bed and still read without having to hold your book in an awkward position.

Buy periscope glasses.

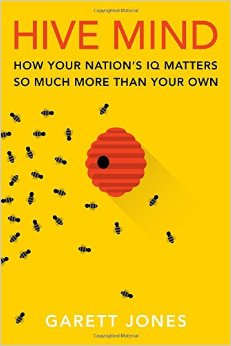
[EDIT: If that wasn’t enough for you there’s also a 2014 Product Recommendations post.]

# Book Review: Hive Mind

[Conflict of interest notice: Author Garett 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 BEEEES!

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:

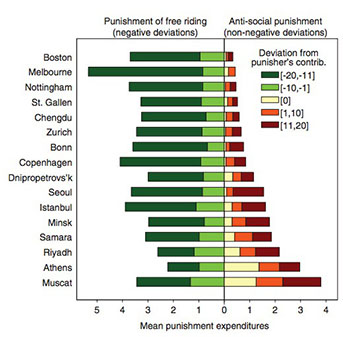
Looking at how individual student test scores predicted those students’ wages later in life, they found that individuals with higher test scores earned only slightly more than average within a given country, but nations with higher average test scores grew expcetionally fast. Here again is the paradox 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 correponded 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 themseles, 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?

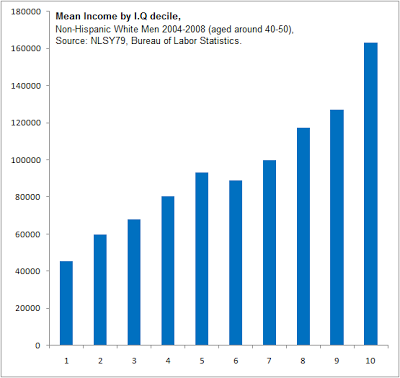
The whole paragraph has the feeling of somebody being dragged over a bed of hot coals, from the insistence on referring to unskilled immigrants as “currently less-skilled” and natives as “so-called native less-skilled workers” to the odd proposal at the end that maybe for some reason less-skilled workers 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 Garett 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 descendents. 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 differnet 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.

# List Of Passages I Highlighted In My Copy Of “Hive Mind”

Even though simple logic said “there will be blood”, there were nevertheless long stretches [of WWI] when all was quiet on the Western front. Why? Axelrod said this was due to the power of tit for tat, which we discussed in Chapter 5. German and French soldiers tacitly created their own unwritten peace treaties: if you don’t shoot at our side, we won’t shoot at yours. And if the higher-ups do force us to fire our artillery, we’ll intentionally aim short of your position, or aim long of it. These rules were never written down, and they were rarely even spoken as far as we can tell – but they left their mark. Take this example. One day the German artillery fired toward the British side without doing any damage. A German infantryman climbed up on a parapet just to deliver an apology to the British: “We are very sorry about that; we hope no one was hurt. It’s not our fault, it’s that damned Prussian artillery” The German infantrymen didn’t want to destroy the fragile truce they had created with their alleged enemies. Military higher-ups hated this tacit cooperation across enemy lines. Fortunately for the officers (but not for the enlisted men) there was a simple solution: move troops around. By swapping one division south two kilometers and another north two kilometers, military officers could turn a repeated prisoner’s dilemma into a literal one-shot game.

This seems almost unbelievable, and absent knowledge of the primary source I can’t be 100% sure, but I guess it fits together with what you always hear about the soldiers celebrating Christmas together and so on.

Some skeptics dismiss IQ tests as just measuring whether you’re good at staring at a piece of paper, coming up with an answer, and writing it down. But the comprehensive IQ test used most often today, the Wechsler mentioned earlier – involves little paper-staring and almost no pencils. The person giving the test (a psychologist or other testing expert) asks you why the seasons change or asks you to recite a list of numbers that she reads out to you. You answer verbally. Later you are handed some wooden puzzle blocks and you try to assemble them into something meaningful.

I hadn’t realized this. There’s a lot of talk about IQ being related to “test-taking skills”, and it’s easy to imagine that people who are used to take paper-and-pencil multiple choice exams might be more comfortable with them, but that’s not what a lot of real IQ tests are and we should stop imagining them that way. This also makes theories about education and the Flynn effect more impressive.

The images are simple shapes like a regular C or a backwards one, and your job is to note, for instance, whether the C is open to the left or the right. So at what point would you no longer be able to do better than random at correctly answering “left” or “right”? When the image flashes for just an eighth of a second? A 32nd? A 128th? That’s the key variable the researcher keeps track of in this study…can it possibly be the case that people who only need to see the image for a tiny fraction of a second tend to have higher average IQ scores than people who need to see it for an eighth or sixteenth of a second? Summarizing “dozens of studies” run on “four continents” dating back to the 1970s, psychologist Ian Deary says, “the overall answer is yes, there is a moderate association between how good people are at the inspection time test and how well they score on intelligence tests.”

…speaking of IQ not just being about test-taking. Jones also goes through some studies on reaction time and reverse digit span, not to mention literal brain size as measured by MRI.

IQ tests do about the same as the best kind of job interviews – structured job interviews in which the interviewer carefully designs the questions beforehand and sticks to the same ones with each candidate – and better than most of the methods people use to choose employees…IQ tests are even better at predicting outcomes when the job requires higher skills. Back in the 1960s, the Bell Telephone System gave its entry level management trainees an IQ-type test along with a number of personality tests…when, after two decades, the company looked back to see which tests did the best job of predicting which trainees eventually rose the highest in the company hierarchy, the IQ-type test did the best job, beating out the personality tests. Looking across many studies of IQ in the elite workforce, one review says “general cognitive ability is the best single predictor of executive/professional-level performance, just as it is of performance in the middle to high end range of the general workforce.”

What I get out of this is “apparently people should structure their job interviews more”.

One might contend that it’s harder to measure social or emotional intelligence than it is to measure more conventional intelligence. But psychologists have tried: they’ve checked to see whether people with more social or emotional intelligence tend to have higher IQs, and so far it looks like they do. The relationship often isn’t as strong as the relationship between, say, a person’s vocabulary test scores and their Raven matrices scores, but the results are clear: IQ scores predict practical social skills.

On the other hand, all of these “measures of social intelligence” are things like “can you read this person’s face to determine what emotion they are feeling?”; I wonder how well that correlates with tasks like “work your way up to becoming coolest person in your high school class” or “win the heart of your love interest”. There is a George Washington Social Intelligence Test, but it doesn’t seem to actually involve becoming the universally beloved father of your country.

But what did the apparently more cautious, more careful Wicherts report? He said that the average IQ in sub-Saharan Africa was about 82 – corresponding to the 12th percentile in the United Kingdom…that’s an improvement from [Lynn’s estimate of] 70, and it’s an improvement that arose partly because Wicherts chose to throw out samples of students who came from families with nutrition problems and low socioeconomic status. The Wicherts average of 82 only includes samples of apparently healthy students from families that have typical socioeconomic status. And in a region of the world with as much poverty and disease as sub-Saharan Africa, that decision is quite likely to leave out substantial portions of the population. To further test the data and get the best average, Wicherts wrote a separate paper that looked at onl the best test samples…Wicherts’ best samples of students have an average IQ score of 76. That’s at the 5th percentile within the United Kingdom.

I keep hearing wildly different numbers for sub-Saharan African IQ, but 76 seems pretty plausible. Note that if the higher 82 number which I keep seeing around is true, it would throw a wrench in a lot of things. Remember that African-Americans are at about 85, so if going from Africa to America only gains you at most 3 IQ points, everything we know about the Flynn Effect falls apart. Even 76 -> 85 is a kind of low Flynn Effect estimate, but I’ll allow it if we speculate that these Africans probably have been at least a little Flynnified, and maybe African Americans are still impoverished enough not to have been fully Flynnified.

Even in rural Pakistan, higher Raven’s IQ predicts higher wages. One might think that thousands of miles away from the Western universities where the tsts were designed, abstract IQ tests would have no power to predict which workers earned more and which earned less – but an IQ test made up of boxes and lines and circles had a modest ability to predict a person’s wages across rural Pakistan, just like in the United States.

Also relevant to IQ tests and culture.

Country-level results go back decades. As early as the 1960s, studies in Taiwan and Hong Kong found average IQ scores slightly above the European average. This happened at a time when these countries’ economies were growing fast, but were still poor by US and UK standards. Any simple story that “wealth causes IQ” has to account for the puzzlingly high average scores found in Taiwan and Hong Kong decades ago…a healthy environment helps to boost IQ, but it can’t be the whole story. The high average IQs of East Asia and Singapore have yet to be fully explained.

See, this is another thing that confuses me about the Flynn Effect. Taiwan etc are still slightly above the European average. So either they had fully caught up to Europe in Flynnification by 1960 (how?!) or Asians have some sort of magic anti-Flynn-effect armor. Didn’t Ron Unz speculate something like that once?

The key evidence comes from the horrific experience of famine in the Netherlands during World War II. Some Dutch towns were cut off from regular food supplies toward the end of the war – daily calories were down to about one third of recommended levels at one point – and the children in gestation during that period had low birth weight. But when these children grew up, their IQs were nearly identical to those of children born a few years earlier or a few years later: massive in utero food deficiencies had no long-term effect on child IQ…famines lasting a few months might not matter for a child’s brain development. But what about chronic malnutrition among the young?…Again the experimental method offers an answer. In one study of Guatemalan villagers, some malnourished children where given a protein supplement and some were given a nonprotein supplement, and then their cognitive skills were measured in few different ways. Of the two interventions, the protein supplement gave a bigger boost to student scores.

I was expecting something awful but on first glance the paper actually looks really good. I’ve uploaded it here if you want to see. They don’t measure the results in IQ so I’m not sure exactly what the magnitude is. It only worked on people who were already poor and probably not well-nourished. Possibly another argument for vegetarians taking protein supplements?

So far it appears that schooling clearly boosts crystallized intelligence, but the typical efect of school on fluid intelligence may be small or even nonexistent…[but] here are two examples of apparently successful increases in fluid intelligence. First, in a study in Sudan, a few months of training on the abacus appeared to boost Raven’s matrix scores dramatically. Second, a study of Israeli schoolchildren, comparing students of similar ages who were born just before or just after the school’s age cutoff: because students born before the cutoff get an extra year of formal schooling, the age cutoff let the researchers identify the effect of an extra year of schooling on student IQ scores. The researchers found a bigger effect on vocabulary scores than on matrix scores, but nevertheless matrix scores indeed rose for students who had received an extra year of schooling.

Right now I am pretty confused on what affects fluid vs. crystalline intelligence. Also, really want to know if the school year results hold up 10 or 20 years later.

In the case of classroom learning, the evidence is mixed on whether your child’s standardized test scores will likely rise if she’s in a class room of high achievers or tend to fall if she’s placed in a classroom of weaker students. This is one question that’s been tested and retested in numerous ways in classrooms around the world, and perhaps the best way to summarize the vast literature is to say that some signs point to positive peer effects and some point to no peer effects. Amid the ambiguous findings, the clearest peer effect is that disruptive students hurt learning.

I could have told you that in third grade.

Sub-Saharan Africa was the last major region of the world to eliminate lead from its gasoline, a goal that apparently was reached in 2006.

That is exciting and will hopefully cause some progress for the region.

Did your neighbor win the lottery and buy a nicer car? That means that you’re more likely to buy a nicer car too. At least that’s what economist Peter Kuhn and his coauthors found when studying the results of lotteries in the Netherlands: people who won big lottery prizes tended to buy nicer cars (no surprise), but people who lived near the winners tended to buy nicer cars too. Buying is a social activity: most of us try at least a little to fit in with our neighbors.

This section was titled “Staying Frugal Like The Joneses”. Stop tooting your own horn, Garrett.

One study had twins play a repeated prisoner’s dilemma game against each other for a hundred rounds, and they found that higher-IQ pairs of twins did cooperate more often than lower-IQ pairs of twins.

I feel like making twins play the prisoners’ dilemma against each other is the sort of thing where you risk accidentally making the machinery of the universe divide by zero and explode.

Here’s the most exciting result from my experiments with Omar and Jaap: on average, over the course of the entire experiment, higher-IQ pairs were five times more cooperative than higher-IQ individuals. The link between IQ and cooperation was an emergent phenomenon: it arose not from smart individual players but from smart pairs of players.

There was a similar result in a dataset taken by looking at all the prisoners’ dilemmas done in different colleges’ psych programs and such, then sorting them by the SAT score of the college. This is a scarier lesson about WEIRD samples than anything I’ve read explicitly on that subject.

So far, two studies in the United Kingdom find that higher [IQ] individuals are more likely to vote, regardless of other things known about the person such as his social class, his education, and some personality traits…however, a study in the United States drawing on three different surveys finds no substantial evidence that IQ predicts voting behavior.

Thanks a lot, Get Out The Vote people.

A team of Notre Dame political scientists ran a series of deliberative polling studies to investigate these questions. They started by surveying participants’ politics attitudes before they were put into a room with other participants. Once they were in the room they were asked to discuss a prescribed set of topics such as the then-ongoing war in Iraq. After the discussion, they were surveyed again. Participants were randomly put into different rooms, some with more Iraq war supporters, some with more opponents, and so on. That made it possible to check whether their post-discussion opinion moved toward the pre-discussin opinion of the average person in their room: it was possible to see if they were conforming. The researchers ran 330 groups of about 10 people each, so if conformity was even a modestly strong force it should have shown up. Here’s what they conclude: “After several years of experimentation, we have found little evidence that group composition influences postdiscussion attitude. This persistent finding, which holds across a range of experimental variations and for subgroups defined by political knowledge, suggests that the expectations of the Asch literature do not apply to the Deliberative Poll setting”

I’m not sure how surprised to be about this. I guess it makes sense that political issues like the Iraq War are so polarizing that you can’t change people’s minds about them with a short discussion. It’s still surprising to find something without an Asch conformity effect at all.

Psychologist Christopher Chabris and his coauthors looked at team efforts another way: they checked to see if there was a “da Vinci effect” for teams, a tendency for teams that did well on one kind of task such as a team game of checkers to do well on another task such as taking an IQ test as a team. Indeed, they found such an efect, which they called a c facor, presumably as homage to the general or g factor of human intelligence. And what predicted which teams did better overall? The strongest two factors were whether team members routinely took turns speaking, and how well team members did on a test of reading the emotions of others…but nevertheless, the IQ of the average group member and the IQ of the highest-scoring group member were weak but still notable predictors of group intelligence, and differences in individual IQ explained over one third of differences in actual team performance.

I guess we should care a lot more about people taking turns speaking. Also, I bet that last statistic is basically entirely dependent on the IQ range in the study.

And I’ll take this space at the bottom here to point out that author Garrett Jones very kindly replied to some of my criticisms in the review here. I reply to his reply here, followed by a long conversation with Pseudoerasmus that continues here. I am not quite to the point of sticking this on the Mistakes page, especially since Arnold Kling had the same concern even before me, but I admit it some of this is more complicated than I originally expected and I might end up going either way on it.

# Should AI Be Open?

I.

H.G. Wells’ 1914 sci-fi book The World Set Free did a pretty good job predicting nuclear weapons:

They did not see it until the atomic bombs burst in their fumbling hands…before the last war began it was a matter of common knowledge that a man could carry about in a handbag an amount of latent energy sufficient to wreck half a city

Wells believed the coming atomic bombs would be so deadly that we would inevitably create a utopian one-world government to prevent them from ever being used. Sorry, Wells. It was a nice thought.

But imagine that in the 1910s and 1920s, some elites had started thinking really seriously along Wellsian lines. They would worry about what might happen when the first nation – let’s say America – got the Bomb. It would be unstoppable in battle and might rule the world with an iron fist. Such a situation would be the end of human freedom and progress.

So in 1920, these elites pooled their resources and made their own Manhattan Project. Their efforts bore fruit, and they learned a lot about nuclear fission; in particular, they learned that uranium was a necessary raw material. The world’s uranium sources were few enough that a single nation or coalition could get a monopoly upon them; the specter of atomic despotism seemed more worrying than ever.

They got their physicists working overtime and discovered a new type of nuke that required no uranium at all. In fact, once you understood the principles you could build one out of parts from a Model T engine. The only downside was that if you didn’t build it exactly right, its usual failure mode was to detonate on the workbench in an uncontrolled hyper-reaction that would blow the entire hemisphere to smithereens.

And so the intellectual and financial elites declared victory – no one country could monopolize atomic weapons now – and sent step-by-step guides to building a Model T nuke to every household in the world. Within a week, both hemispheres were blown to very predictable smithereens.

II.

Some of the top names in Silicon Valley have just announced a new organization, OpenAI, dedicated to “advancing digital intelligence in the way that is most likely to benefit humanity as a whole…as broadly and evenly distributed as possible.” Co-chairs Elon Musk and Sam Altman talk to Steven Levy:

Levy: How did this come about? […]

Musk: Philosophically there’s an important element here: we want AI to be widespread. There’s two schools of thought?—?do you want many AIs, or a small number of AIs? We think probably many is good. And to the degree that you can tie it to an extension of individual human will, that is also good. […]

Altman: We think the best way AI can develop is if it’s about individual empowerment and making humans better, and made freely available to everyone, not a single entity that is a million times more powerful than any human. Because we are not a for-profit company, like a Google, we can focus not on trying to enrich our shareholders, but what we believe is the actual best thing for the future of humanity.

Levy: Couldn’t your stuff in OpenAI surpass human intelligence?

Altman: I expect that it will, but it will just be open source and useable by everyone instead of useable by, say, just Google. Anything the group develops will be available to everyone. If you take it and repurpose it you don’t have to share that. But any of the work that we do will be available to everyone.

Levy: If I’m Dr. Evil and I use it, won’t you be empowering me?

Musk: I think that’s an excellent question and it’s something that we debated quite a bit.

Altman: There are a few different thoughts about this. Just like humans protect against Dr. Evil by the fact that most humans are good, and the collective force of humanity can contain the bad elements, we think its far more likely that many, many AIs, will work to stop the occasional bad actors than the idea that there is a single AI a billion times more powerful than anything else. If that one thing goes off the rails or if Dr. Evil gets that one thing and there is nothing to counteract it, then we’re really in a bad place.

Both sides here keep talking about who is going to “use” the superhuman intelligence a billion times more powerful than humanity, as if it were a microwave or something. Far be it from me to claim to know more than Musk or Altman about anything, but I propose that the correct answer to “what would you do if Dr. Evil used superintelligent AI?” is “cry tears of joy and declare victory”, because anybody at all having a usable level of control over the first superintelligence is so much more than we have any right to expect that I’m prepared to accept the presence of a medical degree and ominous surname.

A more Bostromian view would forget about Dr. Evil, and model AI progress as a race between Dr. Good and Dr. Amoral. Dr. Good is anyone who understands that improperly-designed AI could get out of control and destroy the human race – and who is willing to test and fine-tune his AI however long it takes to be truly confident in its safety. Dr. Amoral is anybody who doesn’t worry about that and who just wants to go forward as quickly as possible in order to be the first one with a finished project. If Dr. Good finishes an AI first, we get a good AI which protects human values. If Dr. Amoral finishes an AI first, we get an AI with no concern for humans that will probably cut short our future.

Dr. Amoral has a clear advantage in this race: building an AI without worrying about its behavior beforehand is faster and easier than building an AI and spending years testing it and making sure its behavior is stable and beneficial. He will win any fair fight. The hope has always been that the fight won’t be fair, because all the smartest AI researchers will realize the stakes and join Dr. Good’s team.

Open-source AI crushes that hope. Suppose Dr. Good and his team discover all the basic principles of AI but wisely hold off on actually instantiating a superintelligence until they can do the necessary testing and safety work. But suppose they also release what they’ve got on the Internet. Dr. Amoral downloads the plans, sticks them in his supercomputer, flips the switch, and then – as Dr. Good himself put it back in 1963 – “the human race has become redundant.”

The decision to make AI findings open source is a tradeoff between risks and benefits. The risk is letting the most careless person in the world determine the speed of AI research – because everyone will always have the option to exploit the full power of existing AI designs, and the most careless person in the world will always be the first one to take it. The benefit is that in a world where intelligence progresses very slowly and AIs are easily controlled, nobody can use their sole possession of the only existing AI to garner too much power.

But what if we don’t live in a world where progress is slow and control is easy?

III.

If AI saunters lazily from infrahuman to human to superhuman, then we’ll probably end up with a lot of more-or-less equally advanced AIs that we can tweak and fine-tune until they cooperate well with us. In this situation, we have to worry about who controls those AIs, and it is here that OpenAI’s model makes the most sense.

But Bostrom et al worry that AI won’t work like this at all. Instead there could be a “hard takeoff”, a subjective discontinuity in the function mapping AI research progress to intelligence as measured in ability-to-get-things-done. If on January 1 you have a toy AI as smart as a cow, and on February 1 it’s proved the Riemann hypothesis and started building a ring around the sun, that was a hard takeoff.

(I won’t have enough space here to really do these arguments justice, so I once again suggest reading Bostrom’s Superintelligence if you haven’t already. For more on what AI researchers themselves think of these ideas, see AI Researchers On AI Risk.)

Why should we expect a hard takeoff? First, it’s happened before. It took evolution twenty million years to go from cows with sharp horns to hominids with sharp spears; it took only a few tens of thousands of years to go from hominids with sharp spears to moderns with nuclear weapons. Almost all of the practically interesting differences in intelligence occur within a tiny window that you could blink and miss.

If you were to invent a sort of objective zoological IQ based on amount of evolutionary work required to reach a certain level, complexity of brain structures, etc, you might put nematodes at 1, cows at 90, chimps at 99, homo erectus at 99.9, and modern humans at 100. The difference between 99.9 and 100 is the difference between “frequently eaten by lions” and “has to pass anti-poaching laws to prevent all lions from being wiped out”.

Worse, the reasons we humans aren’t more intelligent are really stupid. Even people who find the idea abhorrent agree that selectively breeding humans for intelligence would work in some limited sense. Find all the smartest people, make them marry each other for a couple of generations, and you’d get some really smart great-grandchildren. But think about how weird this is! Breeding smart people isn’t doing work, per se. It’s not inventing complex new brain lobes. If you want to get all anthropomorphic about it, you’re just “telling” evolution that intelligence is something it should be selecting for. Heck, that’s all that the African savannah was doing too – the difference between chimps and humans isn’t some brilliant new molecular mechanism, it’s just sticking chimps in an environment where intelligence was selected for so that evolution was incentivized to pull out a few stupid hacks. The hacks seem to be things like “bigger brain size” (did you know that both among species and among individual humans, brain size correlates pretty robustly with intelligence, and that one reason we’re not smarter may be that it’s too hard to squeeze a bigger brain through the birth canal?) If you believe in Greg Cochran’s Ashkenazi IQ hypothesis, just having a culture that valued intelligence on the marriage market was enough to boost IQ 15 points in a couple of centuries, and this is exactly the sort of thing you should expect in a world like ours where intelligence increases are stupidly easy to come by.

I think there’s a certain level of hard engineering/design work that needs to be done for intelligence, a level way below humans, and after that the limits on intelligence are less about novel discoveries and more about tradeoffs like “how much brain can you cram into a head big enough to fit out a birth canal?” or “wouldn’t having faster-growing neurons increase your cancer risk?” Computers are not known for having to fit through birth canals or getting cancer, so it may be that AI researchers only have to develop a few basic principles – let’s say enough to make cow-level intelligence – and after that the road to human intelligence runs through adding the line NumberOfNeuronsSimulated = 100000000000 to the code, and the road to superintelligence runs through adding another zero after that.

(Remember, it took all of human history from Mesopotamia to 19th-century Britain to invent a vehicle that could go as fast as a human. But after that it only took another four years to build one that could go twice as fast as a human.)

If there’s a hard takeoff, OpenAI’s strategy stops being useful. There’s no point in ensuring that everyone has their own AIs, because there’s not much time between the first useful AI and the point at which things get too confusing to model and nobody “has” the AIs at all.

IV.

OpenAI’s strategy also skips over a second aspect of AI risk: the control problem.

All of this talk of “will big corporations use AI?” or “will Dr. Evil use AI?” or “Will AI be used for the good of all?” presuppose that you can use an AI. You can certainly use an AI like the ones in chess-playing computers, but nobody’s very scared of the AIs in chess-playing computers either. What about AIs powerful enough to be scary?

Remember the classic programmers’ complaint: computers always do what you tell them to do instead of what you meant for them to do. Computer programs rarely do what you want the first time you test them. Google Maps has a relatively simple task (plot routes between Point A and Point B), has been perfected over the course of years by the finest engineers at Google, has been ‘playtested’ by tens of millions of people day after day, and still occasionally does awful things like suggest you drive over the edge of a deadly cliff, or tell you to walk across an ocean and back for no reason on your way to the corner store.

Humans have a robust neural architecture, to the point where you can logically prove that what they’re doing is suboptimal and they’ll shrug and say they they’re going to do it anyway. Computers aren’t like this unless we make them so, itself a hard task. They are naturally fragile and oriented toward specific goals. An AI that ended up with a drive as perverse as Google Maps’ occasional tendency to hurl you off cliffs would not be necessarily self-correcting. A smart AI might be able to figure out that humans didn’t mean for it to have the drive it did. But that wouldn’t cause it to change its drive, any more than you can convert a gay person to heterosexuality by patiently explaining to them that evolution probably didn’t mean for them to be gay. Your drives are your drives, whether they are intentional or not.

When Google Maps tells people to drive off cliffs, Google quietly patches the program. AIs that are more powerful than us may not need to accept our patches, and may actively take action to prevent us from patching them. If an alien species showed up in their UFOs, said that they’d created us but made a mistake and actually we were supposed to eat our children, and asked us to line up so they could insert the functioning child-eating gene in us, we would probably go all Independence Day on them; computers with more goal-directed architecture would if anything be even more willing to fight such changes.

If it really is a quick path from cow-level AI to superhuman-level AI, it would be really hard to test the cow-level AI for stability and expect it to stay stable all the way up to superhuman-level – superhumans have a lot more ways to cause trouble than cows do. That means a serious risk of superhuman AIs that want to do the equivalent of hurl us off cliffs, and which are very resistant to us removing that desire from them. We may be able to prevent this, but it would require a lot of deep thought and a lot of careful testing and prodding at the cow-level AIs to make sure they are as prepared as possible for the transition to superhumanity.

And we lose that option by making the AI open source. Make such a program universally available, and while Dr. Good is busy testing and prodding, Dr. Amoral has already downloaded the program, flipped the switch, and away we go.

V.

Once again: The decision to make AI findings open source is a tradeoff between risks and benefits. The risk is that in a world with hard takeoffs and difficult control problems, you get superhuman AIs that hurl everybody off cliffs. The benefit is that in a world with slow takeoffs and no control problems, nobody will be able to use their sole possession of the only existing AI to garner too much power.

But the benefits just aren’t clear enough to justify that level of risk. I’m still not even sure exactly how the OpenAI founders visualize the future they’re trying to prevent. Are AIs fast and dangerous? Are they slow and easily-controlled? Does just one company have them? Several companies? All rich people? Are they a moderate advantage? A huge advantage? None of those possibilities seem dire enough to justify OpenAI’s tradeoff against safety.

Are we worried that AI will be dominated by one company despite becoming necessary for almost every computing application? Microsoft Windows is dominated by one company and became necessary for almost every computing application. For a while people were genuinely terrified that Microsoft would exploit its advantage to become a monopolistic giant that took over the Internet and something something something. Instead, they were caught flat-footed and outcompeted by Apple and Google, plus if you really want you can use something open-source like Linux instead. And new versions of Windows inevitably end up hacked and up on The Pirate Bay anyway.

Or are we worried that AIs will somehow help the rich get richer and the poor get poorer? This is a weird concern to have about a piece of software which can be replicated pretty much for free. Windows and Google Search are both fantastically complex products of millions of man-hours of research; Google is free and Windows comes bundled with your computer. In fact, people have gone through the trouble of creating fantastically complex competitors to both and providing those free of charge, to the point where multiple groups are competing to offer people fantastically complex software for free. While it’s possible that rich people will be able to afford premium AIs, it is hard for me to weigh “rich people get premium versions of things” on the same scale as “human race likely destroyed”. Like, imagine the sort of dystopian world where rich people had nicer things than the rest of us. It’s too horrifying even to contemplate.

Or are we worried that AI will progress really quickly and allow someone to have completely ridiculous amounts of power? But remember, there’s still a government and it tends to look askance on other people becoming powerful enough to compete with it. If some company is monopolizing AI and getting too big, the government will break it up, the same way they kept threatening to break up Microsoft when it was getting too big. If someone tries to use AI to exploit others, the government can pass a complicated regulation against that. You can say a lot of things about the United States government, but you can’t say that they never pass complicated regulations forbidding people from doing things.

Or are we worried that AI will be so powerful that someone armed with AI is stronger than the government? Think about this scenario for a moment. If the government notices someone getting, say, a quarter as powerful as it is, it’ll probably take action. So an AI user isn’t likely to overpower the government unless their AI can become powerful enough to defeat the US military too quickly for the government to notice or respond to. But if AIs can do that, we’re back in the intelligence explosion/fast takeoff world where OpenAI’s assumptions break down. If AIs can go from zero to more-powerful-than-the-US-military in a very short amount of time while still remaining well-behaved, then we actually do have to worry about Dr. Evil and we shouldn’t be giving him all our research.

Or are we worried that some big corporation will make an AI more powerful than the US government in secret? I guess this is sort of scary, but it’s hard to get too excited about. So Google takes over the world? Fine. Do you think Larry Page would be a better or worse ruler than one of these people? What if he had a superintelligent AI helping him, and also everything was post-scarcity? Yeah, I guess all in all I’d prefer constitutional limited government, but this is another supposed horror scenario which doesn’t even weigh on the same scale as “human race likely destroyed”.

If OpenAI wants to trade off the safety of the human race from rogue AIs in order to get better safety against people trying to exploit control over AIs, they need to make a much stronger case than anything I’ve seen so far for why the latter is such a terrible risk.

There was a time when the United States was the only country with nukes. Aside from poor Hiroshima and Nagasaki, it mostly failed to press its advantage, bumbled its way into letting the Russians steal the schematics, and now everyone from Israel to North Korea has nuclear weapons and things are pretty okay. If we’d been so afraid of letting the US government have its brief tactical advantage that we’d given the plans for extremely unstable super-nukes to every library in the country, we probably wouldn’t even be around to regret our skewed priorities.

Elon Musk famously said that AIs are “potentially more dangerous than nukes”. He’s right – so AI probably shouldn’t be open source any more than nukes should.

VI.

And yet Elon Musk is involved in this project. So are Sam Altman and Peter Thiel. So are a bunch of other people who have read Bostrom, who are deeply concerned about AI risk, and who are pretty clued-in.

My biggest hope is that as usual they are smarter than I am and know something I don’t. My second biggest hope is that they are making a simple and uncharacteristic error, because these people don’t let errors go uncorrected for long and if it’s just an error they can change their minds.

But I worry it’s worse than either of those two things. I got a chance to talk to some people involved in the field, and the impression I got was one of a competition that was heating up. Various teams led by various Dr. Amorals are rushing forward more quickly and determinedly than anyone expected at this stage, so much so that it’s unclear how any Dr. Good could expect both to match their pace and to remain as careful as the situation demands. There was always a lurking fear that this would happen. I guess I hoped that everyone involved was smart enough to be good cooperators. I guess I was wrong. Instead we’ve reverted to type and ended up in the classic situation of such intense competition for speed that we need to throw every other value under the bus just to avoid being overtaken.

In this context, the OpenAI project seems more like an act of desperation. Like Dr. Good needing some kind of high-risk, high-reward strategy to push himself ahead and allow at least some amount of safety research to take place. Maybe getting the cooperation of the academic and open-source community will do that. I won’t question the decisions of people smarter and better informed than I am if that’s how their strategy talks worked out. I guess I just have to hope that the OpenAI leaders know what they’re doing, don’t skimp on safety research, and have a process for deciding which results not to share too quickly.

But I am scared that it’s come to this. It suggests that we really and truly do not have what it takes, that we’re just going to blunder our way into extinction because cooperation problems are too hard for us.

I am reminded of what Malcolm Muggeridge wrote as he watched World War II begin:

All this likewise indubitably belonged to history, and would have to be historically assessed; like the Murder of the Innocents, or the Black Death, or the Battle of Paschendaele. But there was something else; a monumental death-wish, an immense destructive force loosed in the world which was going to sweep over everything and everyone, laying them flat, burning, killing, obliterating, until nothing was left…Nor have I from that time ever had the faintest expectation that, in earthly terms, anything could be salvaged; that any earthly battle could be won or earthly solution found. It has all just been sleep-walking to the end of the night.

# How Bad Are Things?

One “advantage” of working in psychiatry is getting a window into an otherwise invisible world of really miserable people.

I work in a wealthy, mostly-white college town consistently ranked one of the best places to live in the country. If there’s anywhere that you might dare hope wasn’t filled to the brim with people living hopeless lives, it would be here. But that hope is not realized. Every day I get to listen to people describe problems that would seem overwrought if they were in a novel, and made-up if they were in a thinkpiece on The Fragmentation Of American Society.

A perfectly average patient will be a 70 year old woman who used to live somewhere else but who moved her a few years ago after her husband died in order to be closer to family. She has some medical condition or other that prevents her from driving or walking around much, and the family she wanted to be closer to have their own issues, so she has no friends within five hundred miles and never leaves her house except to go to doctors’ appointments. She has one son, who is in jail, and one daughter, who married a drug addict. She also has one grandchild, her only remaining joy in the world – but her drug-addict son-in-law uses access to him as a bargaining chip to make her give him money from her rapidly-dwindling retirement account so he can buy drugs. When she can’t cough up enough quickly enough, he bans her from visiting or talking to the grandchild, plus he tells the grandchild it’s her fault. Her retirement savings are rapidly running out and she has no idea what she will do when they’re gone. Probably end up on the street. Also, her dog just died.

If my patients were to read the above paragraph, there are a handful who would sue me for breach of confidentiality, assuming I had just written down their medical history and gotten a couple of details like the number of children wrong. I didn’t. This is a type.

Here’s another. 60 year old guy who was abused as a child, still has visible scars. Ran off at age 15, got a job in a factory, married let’s say a waitress. There was some kind of explosion in his factory, he got PTSD, now he freaks out every time he steps within a hundred meters of a place where manufacturing is going on. Gradually stopped going outside because there were too many scary loud noises, his wife started yelling at him and telling him he was useless, he started beating his wife, put in jail for a year or two for domestic violence, came out, by this point his wife has run off with another man and took everything he owned with her. Moved in with an abusive uncle who is 80 years old and hates his guts, but the uncle needed a caretaker and the guy needed a place to live and they were each other’s only affordable option. Currently lives off disability payments, but the government keeps trying to cut them off, and he keeps having to spend what little he has on a lawyer to prevent them from taking even that away, but half the time he doesn’t make it to his lawyer appointments because he’s too nervous about going outside. Also he has chronic pain. Also he only sleeps two hours a night because of the nightmares, and he’s tired all the time.

(“You have the pill that fixes all of this, right, Doctor? The one they advertised on TV?”)

A while ago I wrote about how strongly we filter for people who are like us intellectually and politically:

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 a bizarre dark matter parallel universe.

Since starting working in psychiatry, I have realized that we also filter for misery. I think a big part of this is sorting by social class. But it’s in a more subtle way than you might think. That first patient, the 70 year old, might on paper have more than the median income if her dead husband’s pension is high enough. I could even imagine the second patient getting a decent payout from his factory and being financially in the clear for a while. It’s more complicated than that – something to do with being the sort of person who ends up in these sorts of situations.

I have three non-mutually exclusive theories for this:

1. The people who come to a psychiatrist are disproportionately the unhappiest and most disturbed. This is obviously true to some degree. But I got the same sort of people when I worked in general medicine and primary care. Even the people who come to a primary care doctor are going to be a little biased towards the sorts of conditions that produce or result from sickness, but people were still much worse off than I thought.

2. My ordinary life shields me from these people. I don’t live in an especially bad neighborhood, so I won’t meet the unhappiest people there. Unhappy people are really depressing, so their lives won’t be covered as much by newspapers and TV. And insofar as they stay in their homes all the time and never come out or talk to anyone else, that in itself is going to prevent me from meeting them.

3. Or maybe many of the people I know are in fact this unhappy, but they never tell anyone except their psychiatrist all of the pieces necessary to put their life story together.

If it were mostly (1), that would be pretty encouraging and mean I’m just biased toward seeing very unlucky people. If it were mostly (2) or (3), that would be pretty bad, and mean everyone else is biased toward not realizing how unlucky everybody else is.

So I made a short script based on the following information:

– About 1% of people are in prison at any given time  
– About 2% of people are on probation, which can actually be really limiting and unpleasant  
– About 1% of people are in nursing homes or hospices  
– About 2% of people have dementia  
– About 20% of people have chronic pain, though this varies widely with the exact survey question, but we are not talking minor aches here. About two-thirds of people with chronic pain describe it as “constant”, and half of people describe it as “unbearable and excruciating”.  
– About 7% of people have depression in any given year  
– About 2% of people are cognitively disabled aka mentally retarded  
– About 1% of people are schizophrenic  
– About 20% of people are on food stamps  
– About 1% of people are wheelchair-bound  
– About 7% of people are alcoholic  
– About 0.5% of people are chronic heroin users  
– About 5% of people are unemployed as per the official definition which includes only those looking for jobs  
– About 3% of people are former workers now receiving disability payments  
– About 1% of people experience domestic violence each year  
– About 10% of people were sexually abused as children, many of whom are still working through the trauma.  
– Difficult to get statistics, but possibly about 20% of people were physically abused as children, likewise.  
– About 9% of people (male and female) have been raped during their lifetime, likewise.

These numbers might be inflated, since I took them from groups working on these problems and those groups have every incentive to make them sound as bad as possible. There’s also a really big problem where a lot of these are conditional upon one another – that is, a person in prison is not also in a nursing home, but a person who is unemployed is far more likely to be on food stamps. This will likely underestimate both the percent of people who have no problems at all, and the percent of people who have multiple problems at once.

Nevertheless, I ran the script twenty times to simulate twenty different people, and here’s what I got (NP stands for “no problems”):

01. Chronic pain  
02. Alcoholic  
03. Chronic pain  
04. NP  
05. NP  
06. Sexually molested as a child + suffering from domestic violence  
07. Unemployed  
08. Alcoholic  
09. NP  
10. NP  
11. NP  
12. Abused as a child  
13. NP  
14. Chronic pain  
15. NP  
16. Abused as a child + unemployed  
17. NP  
18. Alcoholic + on food stamps  
19. NP  
20. Clinically depressed

If the two problems mentioned above haven’t totally thrown off the calculations, this makes me think Psychiatrist-Me is getting a much better window into reality than Normal-Person-Me.

And remember, this doesn’t count all of the problems that don’t fall into easily quantified categories, like “everyone hates them because they’re really ugly and annoying”. It doesn’t count things that I couldn’t find good statistics on, like “had a child die recently”. It doesn’t count things that I would have gotten in trouble for including, like “autistic” or “single mother”. It doesn’t count a lot of things. Consider that the first patient I mentioned – the homebound seventy year old with no friends who’s being extorted by her drug addict son-in-law – would appear on this list as “NP”.

The world is almost certainly a much worse place than any of us want to admit. And that’s before you’ve even left America.

This is part of why I get enraged whenever somebody on Tumblr says “People in Group X need to realize they have it really good”, or “You’re a Group X member, so stop pretending like you have real problems.” The town where I practice psychiatry is mostly white and mostly wealthy. That doesn’t save it. And whenever some online thinkpiece writer laughs about how good people in Group X have it and how hilarious it is that they sometimes complain about their lives, it never fails that I have just gotten home from treating a member of Group X who attempted suicide.

This is also why I am wary whenever people start boasting about how much better we’re doing than back in the bad old days. That precise statement seems to in fact be true. But people have a bad tendency to follow it up with “And so now most people have it pretty good”. I don’t think we have any idea how many people do or don’t have it pretty good. Nobody who hasn’t read polls would intuitively guess that 40-something percent of Americans are young-Earth creationists. How should they know how many people have it pretty good or not?

I think about all of the miserable people in my psychiatric clinic. Then I multiply by ten psychiatrists in my clinic. Then I multiply by ten similarly-sized clinics in my city. Then I multiply by a thousand such cities in the United States. Then I multiply by hundreds of countries in the world, and by that time my brain has mercifully stopped being able to visualize what that signifies.

This wasn’t supposed to be a Christmas post, but it took me longer than I expected to write, so here we are.

And this wasn’t supposed to be advocating any particular response, but I was recently asked to plug Giving What We Can’s pledge drive, and maybe one of the good responses to realizing how awful things are is committing to donate a little bit of what you’ve got to making them better.

# Things That Are Not Superintelligences

[Not the most interesting topic in the world, but I’m posting it so I have something to link to next time I see this argument]

I talk about superintelligence a lot, usually in the context of AI or genetically engineered humans. And lately I have run into people who say: “But superintelligence already exists! It’s corporations / bureaucracies / teams / civilizations / mind-mapping software” (examples: 1, 2, 3, 4). Sometimes these people go so far as to say these things are in fact superintelligent AIs, since they are technically “artificial”.

No.

Some of these things may be poetically or metaphorically like a superintelligence, in the same way that, I don’t know, the devastation of traditional cultures by modernity is poetically or metaphorically like nuclear war, or whatever. But if every time somebody is trying to talk about nuclear disarmament, other people interject with “But we’ve already had a nuclear war – it’s the nuclear war in the heart of all mankind“, this doesn’t really add to the conversation. In the same way, talking about these metaphorical superintelligences is not a helpful contribution to discussion of literal superintelligences.

Why do I think that there is an important distinction between these kind of collective intelligences and genuine superintelligence?

There is no number of chimpanzees which, when organized into a team, will become smart enough to learn to write.

There is no number of ordinary eight-year-olds who, when organized into a team, will become smart enough to beat a grandmaster in chess.

There is no number of ordinary IQ 90 construction workers who, when organized into a team, will become smart enough to design and launch a space probe to one of the moons of Jupiter.

There is no number of “average” mathematics PhDs who, when organized into a team, will become smart enough to come up with the brilliant groundbreaking proofs of a Gauss or a Ramanujan.

Teams / corporations / cultures have a lot of advantages over individuals. They can use writing and record-keeping to have much better “memories”. They can use computers to be able to calculate and retrieve information more quickly. They can pool their advantages, so that if one person is good at writing and another person good at illustration, they can produce a well-written and beautifully-illustrated book. They can formalize their decision-making processes to route around various biases and react consistently to predictable situations. These are all really good things to be able to do, and it’s why in fact groups of people have outperformed individuals in fields as diverse as “making nuclear bombs” and “coordinating air traffic”.

But there is some aspect of intelligence that they can’t simulate, in which they are forever doomed to be only as smart as their smartest member (if that!). It’s hard to put my finger on exactly, but it seems to have something to do with creative problem-solving ability. A team of people smart enough to solve problems up to Level N may be able to work in parallel to solve many more Level N problems in a given amount of time. But it won’t be able to solve Level N+1 problems. This is why it’s still occasionally useful to have mathematical geniuses around, instead of taking ten average mathematicians and telling them to work together. And unfortunately, this aspect of intelligence is the bottleneck for lots of interesting things like new inventions, proofs, and discoveries.

Further, teams themselves need intelligent people to run in an intelligent way. Steve Jobs led Apple to success by being really really good at marketing. Apple couldn’t have gotten the same results by firing him and replacing him with a marketing department of a hundred low-level employees who had graduated from second-tier marketing programs. This is true not only at the Steve Jobs level but at every level – at some point a Sales Department needs to have good salespeople, not just many well-organized mediocre salespeople. I’m not denying that many well-organized mediocre salespeople can do way better than a few poorly-organized mediocre salespeople, just that you can’t fully route around the need for actually intelligent people.

And finally, teams have a lot of contingent disadvantages over an individual. They work vastly more slowly. Their various parts tend not to know what the other parts are doing. If dictatorial in structure, they fall prey to failures of information; if non-dictatorial, to failures of coordination. Imagine an individual human whose inner soul had Democratic and Republican parties that were constantly trying to sabotage each other, so that if the Democratic part of her got a job interview, the Republican part would immediately try to sabotage the job interview to prevent the Democrats from looking good. Such a person would either be insane or at the very least not get too many jobs.

While it’s possible for improvements in organizational technology to ameliorate some of these contingent problems, so far they generally haven’t: the US government is as dysfunctional as ever, and a lot of corporations are little better. And even if all of the contingent problems were magically solved, that still leaves the fundamental problem where no organization of chimpanzees will ever write a novel.

If we were to actually get superintelligence, that would be a completely different class of entity than another government or corporation. It would also have all of the advantages of these things – arbitrarily much parallel processing ability, arbitrarily much memory, arbitrarily many computational resources – without the disadvantages, and with higher genuine “intelligence” as in problem-solving ability.

I think it’s useful to have a word for this completely different class of things, and that word is “superintelligence”. Teams, corporations, and cultures can use words we already have, like “groups”.

[EDIT: I keep getting the same objection in the comments: if we made a bunch of ordinary eight-year olds follow a simple set of operations that corresponded to a logic gate, and arranged them so that they simulated the structure of Deep Blue, then they could win high-level chess games. This is true. But eight-year-olds could not come up with and implement this idea. A brilliant computer programmer might be able to, but once you’re a brilliant computer programmer, you might as well just build the darned computer instead of implementing it on eight year olds. And any computer programmer so brilliant that they could build a true superintelligence out of eight year olds could build a true superintelligence out of normal computers too. And the same is true of the objection “doesn’t this mean that no amount of stupid neurons could combine into a smart human brain?” Yes, evolution can play the role of the brilliant computer programmer and turn neurons into a working brain. But it’s the organizer – whether that organizer is a brilliant human programmer or an evolutionary process – who is actually doing the work. That “neurons can combine to form a brain” is no more profound than that “transistors can combine to form an AI” – in both cases, it’s the outside organizer doing all the meaningful work. For a really interesting science-fiction treatment of what it would actually mean to implement a superintelligence in human social interaction, read Karl Schroeder’s Lady of Mazes]

[EDIT 2: Advocate of a “super-individual-human-intelligence” project responds with a clarification of terms.]

[EDIT 3: Several people point out in the comments that chess champion Garry Kasparov once played a game of chess against “the world”. Kasparov moved the white pieces, and the black moves were decided by popular vote on a website where various other grandmasters and chess buffs worked together to devise the best strategies. The game was very closely fought and suffered from several irregularities, but Kasparov ended up winning.]

# Introducing Unsong

I.

In retrospect, there had been omens and portents.

(“We are now approaching lunar sunrise,” said William Anders, “and for all the people back on Earth, the crew of Apollo 8 has a message that we would like to send to you.”)

Rivers flowed uphill. A new star was seen in the night sky. A butchered pig was found to have the word “OMEN” written on its liver in clearly visible letters.

(“In the beginning, God created the heaven and the earth. And the earth was without form, and void; and darkness was upon the face of the deep.”)

Lightning struck in clear weather. Toads fell from the clouds. All ten thousand lakes in Minnesota turned to blood; scientists blamed “phytoplankton”.

(“And the Spirit of God moved upon the face of the waters. And God said, Let there be light: and there was light. And God saw the light, that it was good: and God divided the light from the darkness.”)

A majestic golden eagle flew onto the Vatican balcony as Pope Paul VI was saying Mass. The bird gingerly removed the Pontiff’s glasses with its beak, then poked out his left eye before flying away with an awful shriek.

(“And God called the light Day,” said Jim Lovell, “and the darkness He called Night. And the evening and the morning were the first day.”)

A beached whale was found hundreds of miles inland. A baby was born with four eyes.

(“And God said, Let there be a firmament in the midst of the waters, and let it divide the waters from the waters.”)

Pieces of paper with the word “OMEN” written on them fell from the clouds. A beached whale was seen in the night sky. Babies left unattended began to roll slowly, but unmistakeably, uphill.

(“And God made the firmament, and divided the waters which were under the firmament from the waters which were above the firmament: and it was so. And God called the firmament Heaven. And the evening and the morning were the second day.”)

One of the additional eyes on the four-eyed baby was discovered to be the left eye of Pope Paul VI, missing since the eagle incident. The provenance of the fourth eye was never determined.

(“And God said, Let the waters under the heaven be gathered together unto one place,” said Frank Borman, “and let the dry land appear: and it was so.”)

A series of very precise lightning strikes seared the word “OMEN” into the rust-red sand of the Sonora Desert; scientists blamed “phytoplankton”.

(“And God called the dry land Earth; and the gathering together of the waters called he Seas: and God saw that it was good.”)

The New York Stock Exchange rose by perfect integer amounts eleven days in a row. An obstetrician published an article in an obscure medical journal claiming that the kicks of unborn children, interpreted as Morse Code, formed unspeakable and blood-curdling messages.

(“And from the crew of Apollo 8, we close with good night, good luck, a Merry Christmas – and God bless all of – ” [sudden burst of static, then silence])

II.

If I had to choose a high point for the history of the human race thus far, it would be December 24, 1968.

1968 had been a year of shattered dreams. Martin Luther King was murdered in April. Democratic golden boy Robert Kennedy was murdered in June. Soviet tanks crushed the Prague Spring in August. It felt like each spark of hope for a better world was being snuffed out, methodically, one by one.

Then almost without warning, Americans turned on their televisions and learned that a spaceship was flying to the moon. On December 22, the craft beamed a live TV broadcast to Earth informing viewers that they were about to become the first humans ever to approach another celestial body. Communications issues limited the transmission to seventeen minutes, but the astronauts promised a second installment from lunar orbit.

On December 24, 1968, one billion people – more than for any television program before or after in the history of mankind – tuned in for Apollo 8’s short broadcast. The astronauts were half-asleep, frazzled with days of complicated calculations and near-disasters – but their voices were powerful and lucid through the static. Commander Frank Borman introduced the two other members of the crew. They described the moon, as seen up close. “A vast, lonely, forbidding expanse of nothing”. “A very foreboding horizon, a rather dark and unappetizing looking place”. Then the Earth, as seen from afar. “A green oasis, in the big vastness of space.”

Two minutes left till lunar sunrise broke the connection. The astronauts’ only orders from NASA had been to “do something appropriate”

“In the beginning,” read Bill Anders, “God created the heaven and the earth. And the earth was without form, and void; and darkness was upon the face of the deep.”

So for two minutes on Christmas Eve, while a billion people listened, three astronauts read the Book of Genesis from a tiny metal can a hundred miles above the surface of the moon.

Then, mid-sentence, they crashed into the crystal sphere surrounding the world, because it turned out there were far fewer things in Heaven and Earth than were dreamt of in almost anyone’s philosophy.

Everyone has been so kind and encouraging about my short stories that I’m ready to try writing some longer fiction. You can follow along at unsongbook.com, where I’ll be posting new chapters every Sunday and new interludes some Wednesdays. Right now it’s just this prologue and an option to subscribe by email to future updates, but I’ll have the first chapter up by Sunday, January 3.

This is going to be fun.

# 2015 Predictions: Calibration Results

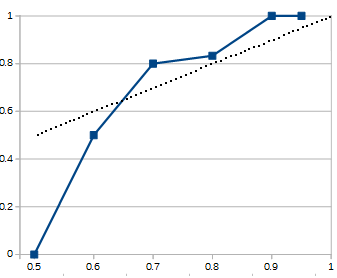
Many people are making predictions for the new year around now. I see fewer people going back and grading the accuracy of their predictions about last year, even though that obviously has a lot of relevance for how seriously we should take this year’s predictions.

In keeping with the SSC tradition, here are the results for my (late) predictions for 2015. Successful predictions are normal-looking, failed predictions are crossed out:

World Events  
1. US will not get involved in any new major war with death toll of > 100 US soldiers: 70%  
2. North Korea’s government will survive the year without large civil war/revolt: 95%  
3. Greece will not announce it’s leaving the Euro: 60%  
3. Neither Russia nor Qatar will lose their World Cups: 80%  
4. Ebola will kill fewer people in second half of 2015 than the in first half: 95%  
5. No terrorist attack in the USA will kill > 100 people: 90%  
6. Assad will remain President of Syria: 70%  
7. Israel will not get in a large-scale war (ie >100 Israeli deaths) with any Arab state: 90%  
8. Syria’s civil war will not end this year: 80%  
9. ISIS will control less territory than it does right now: 70%  
10. ISIS will continue to exist: 80%  
11. Iran will reach a deal with the West on nuclear weapons: 80%  
12. No major civil war in Middle Eastern country not currently experiencing a major civil war: 90%  
13. Iraq’s situation not to get any worse (eg gov’t collapse, new rebellion): 60%  
14. Obamacare will survive the year mostly intact: 60%  
15. Hillary Clinton will be the top-polling Democratic Presidential candidate: 95%  
16. Jeb Bush will be the top-polling Republican candidate: 50%  
17. Trans-Pacific Partnership to pass at least mostly intact: 60%  
18. US official unemployment rate will be less than 7% in Dec 2015: 95%  
19. Bitcoin will end the year higher than $200: 95%  
20. Oil will end the year greater than $60 a barrel: 50%

Personal Life  
21. SSC will remain active: 95%  
22. SSC will get fewer hits in the second half of 2015 than the first half: 60%  
23. At least one SSC post in the second half of 2015 will get > 100,000 hits: 70%  
24. Shireroth will remain active: 90%  
25. I will remain at my same job through the end of 2015: 95%  
26. There will be no further ramifications or lawsuits from either side over the flooding of my house: 80%  
27. I will reach my savings target: 90%  
28. I will get a score at >95th percentile for my year on PRITE: 50% (unknown, haven’t gotten score back)  
29. I will be involved in at least one published/accepted-to-publish research paper by the end of 2015: 60%  
30. I will not break up with any of my current girlfriends: 80%  
31. I will not get any new girlfriends: 50%  
32. I will not finish [project]: 60%  
33. I will attend NYC Solstice ritual: 80%  
34. I will flake out of my plan to lead some kind of Solstice Ritual myself: 60%  
35. I will be living in the house I’m currently trying to arrange to rent: 70%

Scoring  
Of items I marked as 50% confident, 0 were right and 3 were wrong  
Of items I marked as 60% confident, 4 were right and 4 were wrong  
Of items I marked as 70% confident, 4 were right and 1 was wrong  
Of items I marked as 80% confident, 5 were right and 1 was wrong  
Of items I marked as 90% confident, 4 were right and 0 were wrong  
Of items I marked as 95% confident, 7 were right and 0 were wrong



As usual, the dotted line represents perfect calibration; the closer my blue line comes to that, the better I’m doing.

The graph looks like there’s a massive failure at 50%, but this is just an artifact of very few questions at that level. If I’d gotten just one more right, I would be at 33%, ie as close to 50% as it’s possible to get with a set of three. Given that the difference between total success and total failure was just one question, I don’t feel too bad about total failure. Everything else looks pretty good. I’m prepared to call this another successful year.

A side note: Scott Adams has also graded his predictions from the past year, and reports incredible success: 9/9 correct despite going way out on a limb and saying things everyone else found really unlikely (like that Trump would stay Republican front-runner). The obvious way to accomplish like that is to make lots of things that vaguely sound like predictions, then only highlight and count the ones that end up correct; after a quick scan of Adams’ blog, there’s no sign that he’s doing this; his win seems pretty genuine. Another method might be to make vague predictions and grade them in your favor, and there is some sign of this – for example, someone going off Clinton’s poll numbers versus Sanders could say they are in fact still going up. Nevertheless, it’s obviously been a good year for Adams, and I’d be fascinated to see him make a list of official concrete predictions for 2016, all in one place, maybe even associated with confidence levels. [EDIT: Someone else’s more pessimistic analysis; this makes it more pressing that he do everything beforehand this time]