Best of LessWrong: August 2014

1. Why the tails come apart
2. Roles are Martial Arts for Agency
3. [LINK] Speed superintelligence?
4. LW client-side comment improvements
5. Fighting Biases and Bad Habits like Boggarts
6. Quantified Risks of Gay Male Sex
7. Six Plausible Meta-Ethical Alternatives
8. Hal Finney has just died.
9. Multiple Factor Explanations Should Not Appear One-Sided

Best of LessWrong: August 2014

1. Why the tails come apart
2. Roles are Martial Arts for Agency
3. [LINK] Speed superintelligence?
4. LW client-side comment improvements
5. Fighting Biases and Bad Habits like Boggarts
6. Quantified Risks of Gay Male Sex
7. Six Plausible Meta-Ethical Alternatives
8. Hal Finney has just died.
9. Multiple Factor Explanations Should Not Appear One-Sided

Why the tails come apart

[I'm unsure how much this rehashes things 'everyone knows already' - if old hat, feel free to downvote into oblivion. My other motivation for the cross-post is the hope it might catch the interest of someone with a stronger mathematical background who could make this line of argument more robust]

[Edit 2014/11/14: mainly adjustments and rewording in light of the many helpful comments below (thanks!). I've also added a geometric explanation.]

Many outcomes of interest have pretty good predictors. It seems that height correlates to performance in basketball (the average height in the NBA is around 6'7"). Faster serves in tennis improve one's likelihood of winning. IQ scores are known to predict a slew of factors, from income, to chance of being imprisoned, to lifespan.

What's interesting is what happens to these relationships 'out on the tail': extreme outliers of a given predictor are seldom similarly extreme outliers on the outcome it predicts, and vice versa. Although 6'7" is very tall, it lies within a couple of standard deviations of the median US adult male height - there are many thousands of US men taller than the average NBA player, yet are not in the NBA. Although elite tennis players have very fast serves, if you look at the players serving the fastest serves ever recorded, they aren't the very best players of their time. It is harder to look at the IQ case due to test ceilings, but again there seems to be some divergence near the top: the very highest earners tend to be very smart, but their intelligence is not in step with their income (their cognitive ability is around +3 to +4 SD above the mean, yet their wealth is much higher than this) (1).

The trend seems to be that even when two factors are correlated, their tails diverge: the fastest servers are good tennis players, but not the very best (and the very best players serve fast, but not the very fastest); the very richest tend to be smart, but not the very smartest (and vice versa). Why?

Too much of a good thing?

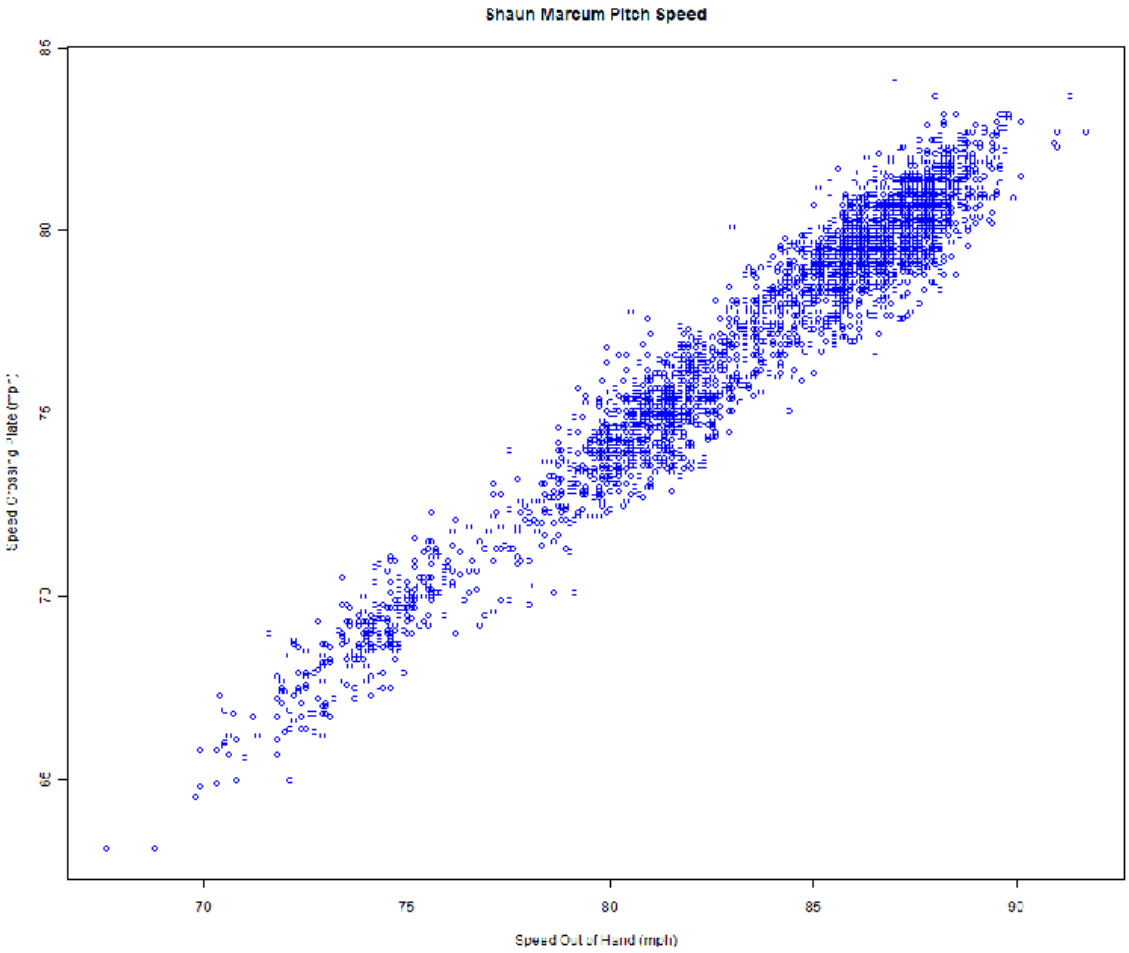
One candidate explanation would be that more isn't always better, and the correlations one gets looking at the whole population doesn't capture a reversal at the right tail. Maybe being taller at basketball is good *up to a point*, but being really tall leads to greater costs in terms of things like agility. Maybe although having a faster serve is better all things being equal, but focusing too heavily on one's serve counterproductively neglects other areas of one's game. Maybe a high IQ is good for earning money, but a stratospherically high IQ has an increased risk of productivity-reducing mental illness. Or something along those lines.

I would guess that these sorts of 'hidden trade-offs' are common. But, the 'divergence of tails' seems pretty ubiquitous (the tallest aren't the heaviest, the smartest parents don't have the smartest children, the fastest runners aren't the best footballers, etc. etc.), and it would be weird if there was always a 'too much of a good thing' story to be told for all of these associations. I think there is a more general explanation.

The simple graphical explanation

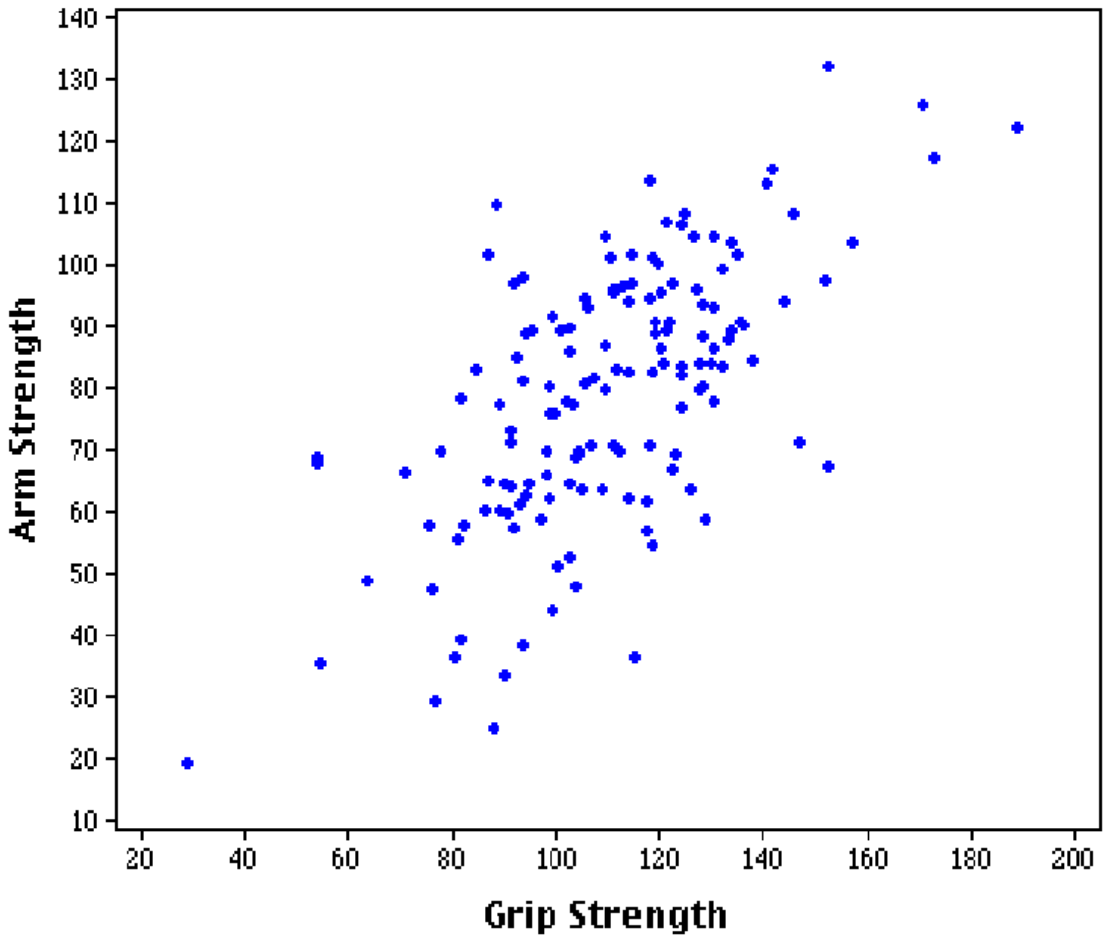
[Inspired by this essay from Grady Towers]

Suppose you make a scatter plot of two correlated variables. Here's one I grabbed off google, comparing the speed of a ball out of a baseball pitchers hand compared to its speed crossing crossing the plate:

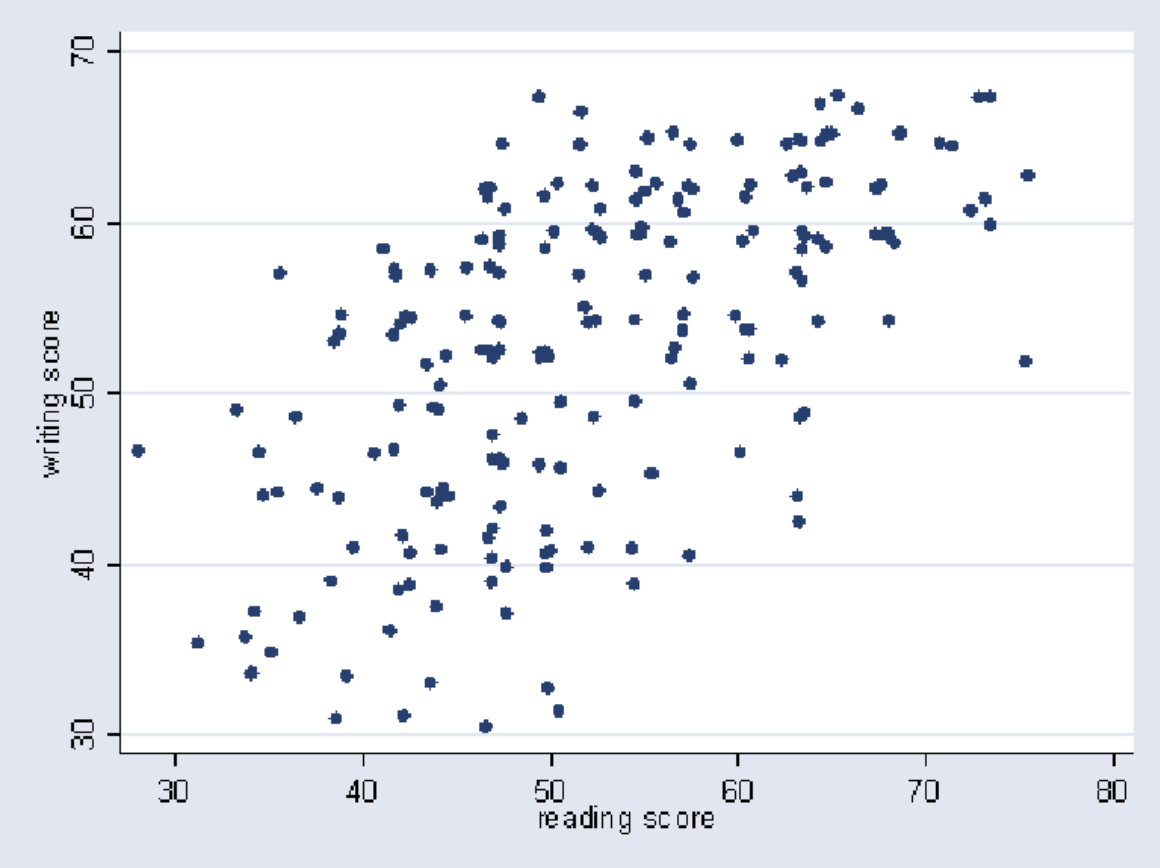


It is unsurprising to see these are correlated (I'd guess the R-square is > 0.8). But if one looks at the extreme end of the graph, the very fastest balls out of the hand *aren't* the very fastest balls crossing the plate, and vice versa. This feature is general. Look at this data (again convenience sampled from googling 'scatter plot') of this:

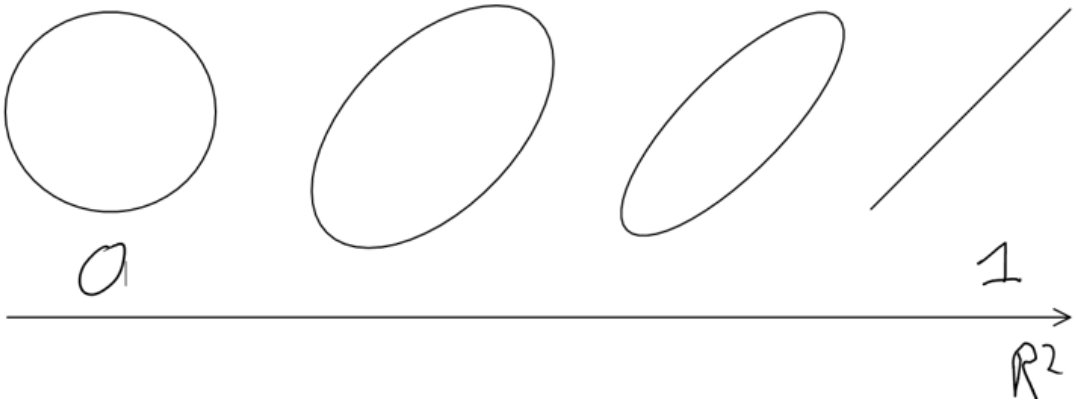
Or this:


Arm Strength
Grip Strength
140
130
120
110
100
90
80
70
60
50
40
30
20
10
20
40
60
80
100
120
140
160
180
200

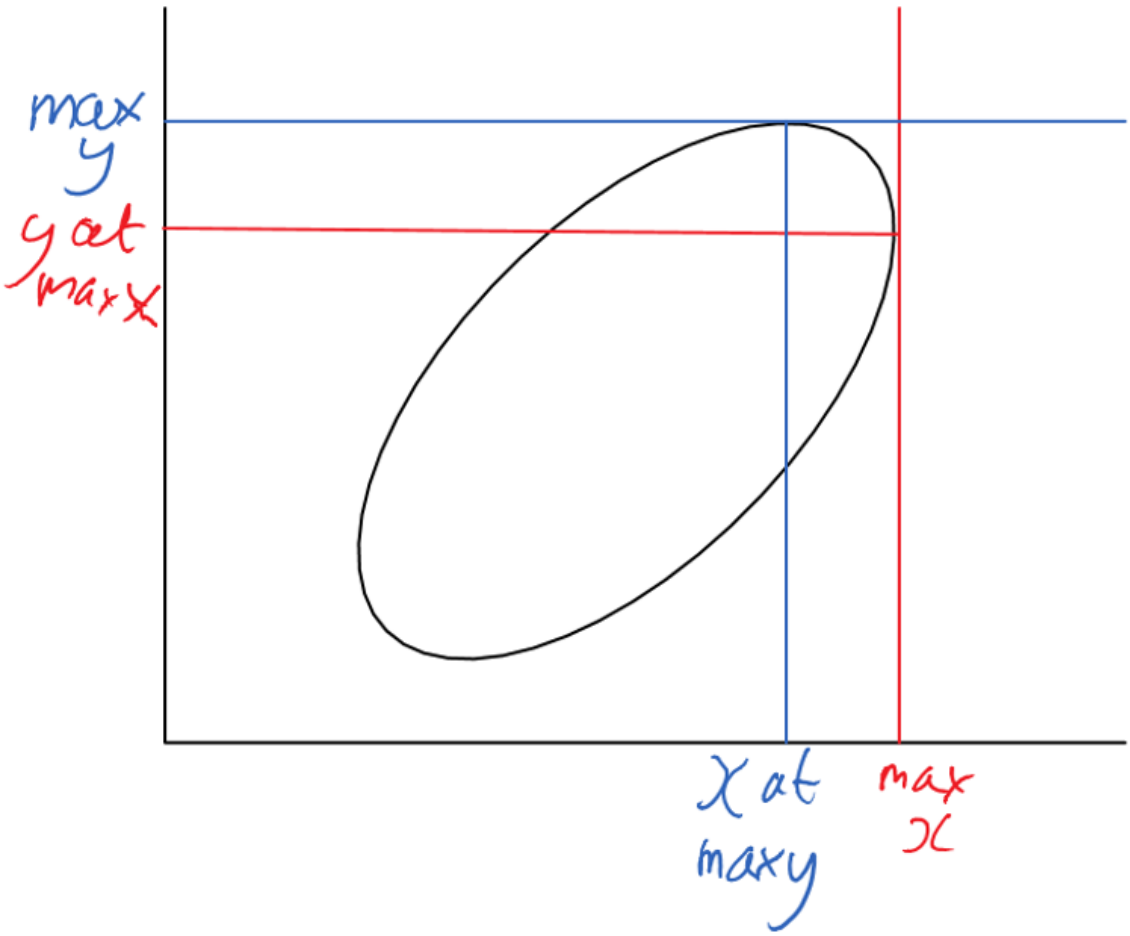
Or this:



Given a correlation, the envelope of the distribution should form some sort of *ellipse*, narrower as the correlation goes stronger, and more circular as it gets weaker: (2)



The thing is, as one approaches the far corners of this ellipse, we see 'divergence of the tails': as the ellipse doesn't sharpen to a point, there are bulges where the maximum x and y values lie with sub-maximal y and x values respectively:



So this offers an explanation why divergence at the tails is ubiquitous. Providing the sample size is largeish, and the correlation not too tight (the tighter the correlation, the larger the sample size required), one will observe the ellipses with the bulging sides of the distribution. (3)

Hence the very best basketball players aren't the very tallest (and vice versa), the very wealthiest not the very smartest, and so on and so forth for any correlated X and Y. If X and Y are "Estimated effect size" and "Actual effect size", or "Performance at T", and "Performance at T+n", then you have a graphical display of winner's curse and regression to the mean.

An intuitive explanation of the graphical explanation

It would be nice to have an intuitive handle on *why* this happens, even if we can be convinced *that* it happens. Here's my offer towards an explanation:

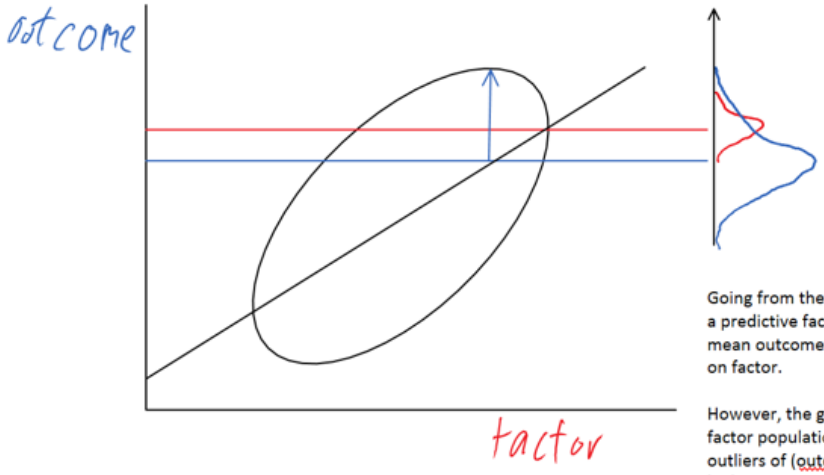
The fact that a correlation is less than 1 implies that *other things matter* to an outcome of interest. Although being tall matters for being good at basketball, strength, agility, hand-eye-coordination matter as well (to name but a few). The same applies to other outcomes where multiple factors play a role: being smart helps in getting rich, but so does being hard working, being lucky, and so on.

For a toy model, pretend that wealth is wholly explained by two factors: intelligence and conscientiousness. Let's also say these are equally important to the outcome, independent of one another and are normally distributed. (4) So, *ceteris paribus*, being more intelligent will make one richer, and the toy model stipulates there aren't 'hidden trade-offs': there's no negative correlation between intelligence and conscientiousness, even at the extremes. Yet the graphical explanation suggests we should still see divergence of the tails: the very smartest shouldn't be the very richest.

The intuitive explanation would go like this: start at the extreme tail - +4SD above the mean for intelligence, say. Although this gives them a massive boost to their wealth, we'd expect them to be average with respect to conscientiousness (we've stipulated they're independent). Further, as this ultra-smart population is small, we'd expect them to fall close to the average in this other independent factor: with 10 people at +4SD, you wouldn't expect any of them to be +2SD in conscientiousness.

Move down the tail to less extremely smart people - +3SD say. These people don't get such a boost to their wealth from their intelligence, but there should be a lot more of them (if 10 at +4SD, around 500 at +3SD), this means one should expect more variation in conscientiousness - it is much less surprising to find someone +3SD in intelligence *and also* +2SD in conscientiousness, and in the world where these things were equally important, they would 'beat' someone +4SD in intelligence but average in conscientiousness. Although a +4SD intelligence person will likely be better than a given +3SD intelligence person (the mean conscientiousness in both populations is 0SD, and so the average wealth of the +4SD intelligence population is 1SD higher than the 3SD intelligence people), the wealthiest of the +4SDs will not be as good as the best of the much larger number of +3SDs. The same sort of story emerges when we look at larger numbers of factors, and in cases where the factors contribute unequally to the outcome of interest.

When looking at a factor known to be predictive of an outcome, the largest outcome values will occur with sub-maximal factor values, as the larger population increases the chances of 'getting lucky' with the other factors:



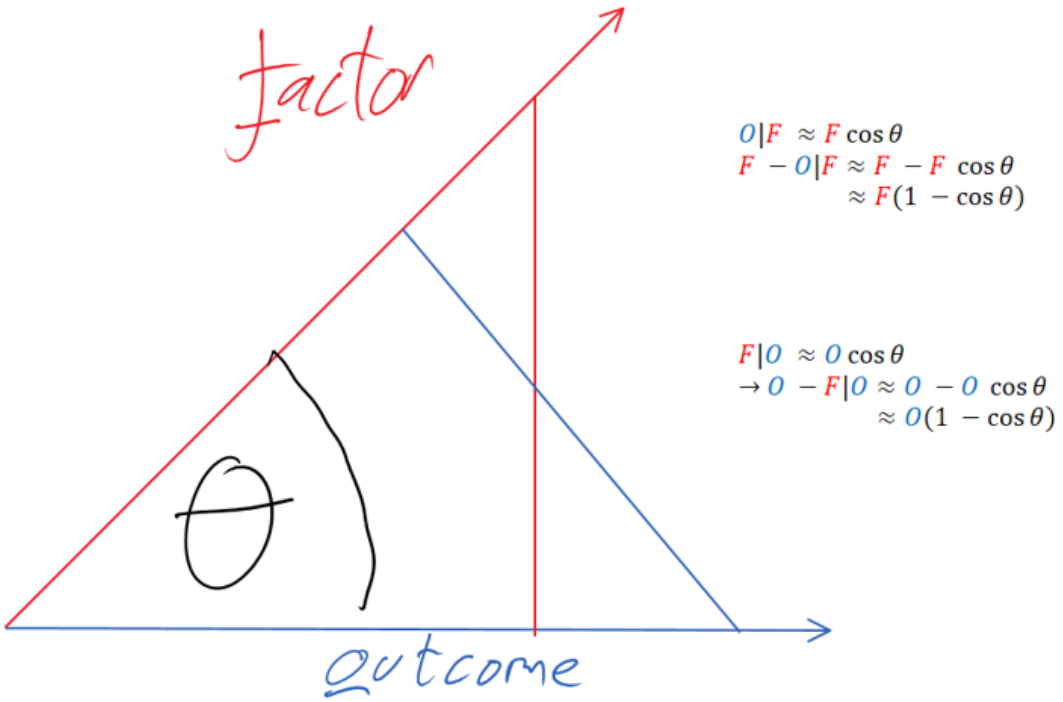
Going from the maximum to sub-maximal values of a predictive factor gives a predictable reduction of mean outcome along the regression line of outcome on factor.

However, the greater population of the sub-maximal factor population gives expectedly more extreme outliers of (outcome|factor), which means the highest outcome for the sub-maximal factor population is higher than the highest outcome for the maximal factor population (right panel).

So that's why the tails diverge.

A parallel geometric explanation

There's also a geometric explanation. The R-square measure of correlation between two sets of data is the same as the cosine of the angle between them when presented as vectors in N-dimensional space (explanations, derivations, and elaborations here, here, and here). (5) So here's another intuitive handle for tail divergence:



Grant a factor correlated with an outcome, which we represent with two vectors at an angle theta, the inverse cosine equal the R-squared. 'Reading off the expected outcome given a factor score is just moving along the factor vector and multiplying by cosine theta to get the distance along the outcome vector. As cos theta is never greater than 1, we see regression to the mean. The geometrical analogue to the tails coming apart is the absolute *difference* in length along factor versus length along outcome|factor scales with the length along the factor; the gap between extreme values of a factor and the less extreme values of the outcome grows linearly as the factor value gets more extreme. For concreteness (and granting normality), an R-square of 0.5 (corresponding to an angle of sixty degrees) means that +4SD (~1/15000) on a factor will be expected to be 'merely' +2SD (~1/40) in the outcome - and an R-square of 0.5 is remarkably strong in the social sciences, implying it accounts for half the variance.(6) The reverse - extreme outliers on outcome are not expected to be so extreme an outlier on a given contributing factor - follows by symmetry.

Endnote: EA relevance

I think this is interesting in and of itself, but it has relevance to Effective Altruism, given it generally focuses on the right tail of various things (What are the *most* effective charities? What is the *best* career? etc.) It generally vindicates worries about regression to the mean or winner's curse, and suggests that these will be pretty insoluble in all cases where the populations are large: even if you have really good means of assessing the best charities or the best careers so that your assessments correlate really strongly with what ones actually

are the best, the very best ones you identify are unlikely to be *actually* the very best, as the tails will diverge.

This probably has limited practical relevance. Although you might expect that *one* of the 'not estimated as the very best' charities is in fact better than your estimated-to-be-best charity, you don't know which one, and your best bet remains your estimate (in the same way - at least in the toy model above - you should bet a 6'11" person is better at basketball than someone who is 6'4".)

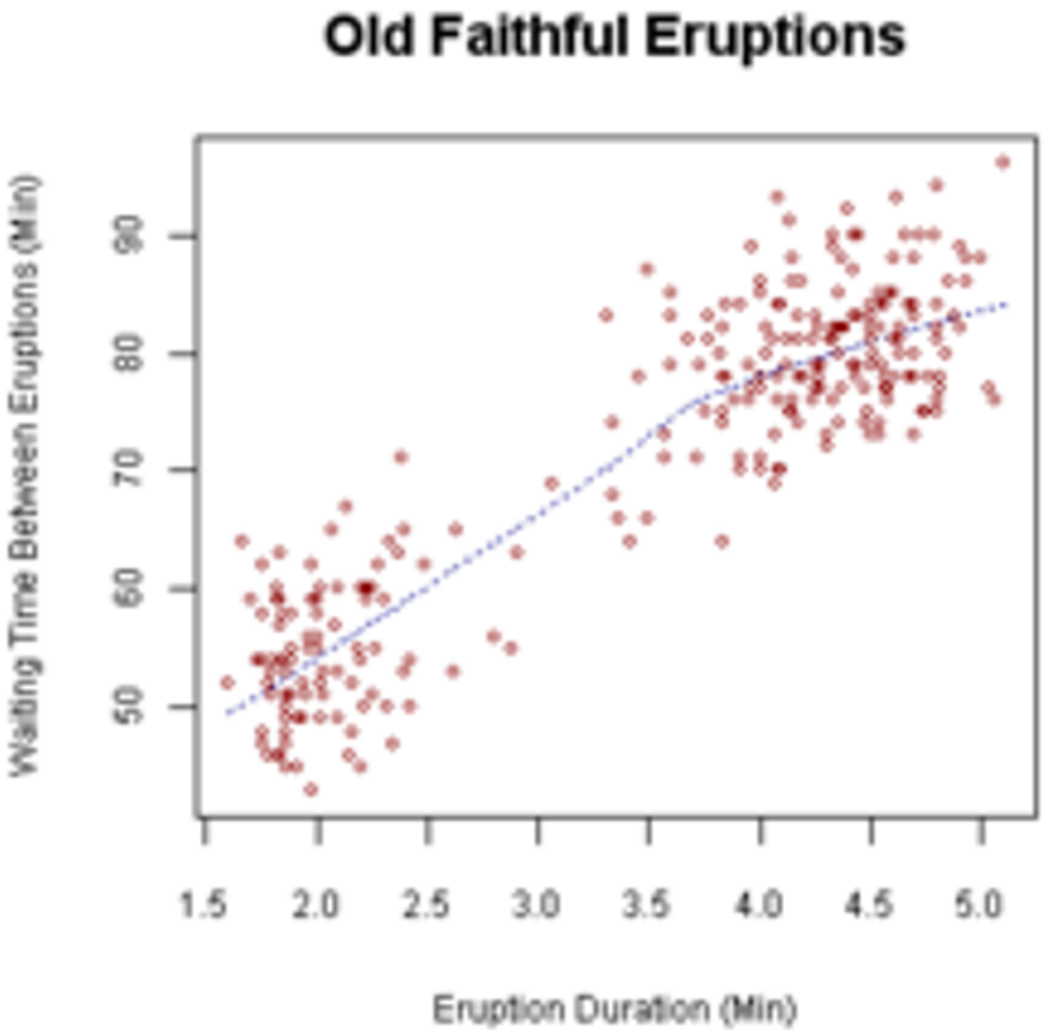
There may be spread betting or portfolio scenarios where this factor comes into play - perhaps instead of funding AMF to diminishing returns when its marginal effectiveness dips below charity #2, we should be willing to spread funds sooner.(6) Mainly, though, it should lead us to be less self-confident.

---

1. Given income isn't normally distributed, using SDs might be misleading. But non-parametric ranking to get a similar picture: if Bill Gates is ~+4SD in intelligence, despite being the richest man in america, he is 'merely' in the smartest tens of thousands. Looking the other way, one might look at the generally modest achievements of people in high-IQ societies, but there are worries about adverse selection.

2. As nshepperd notes below, this depends on something like multivariate CLT. I'm pretty sure this can be weakened: all that is needed, by the lights of my graphical intuition, is that the envelope be *concave*. It is also worth clarifying the 'envelope' is only meant to illustrate the shape of the distribution, rather than some boundary that contains the entire probability density: as suggested by homunq: it is an 'pdf isobar' where probability density is higher inside the line than outside it.

3. One needs a large enough sample to 'fill in' the elliptical population density envelope, and the tighter the correlation, the larger the sample needed to fill in the sub-maximal bulges. The old faithful case is an example where actually you do get a 'point', although it is likely an outlier.



4. It's clear that this model is fairly easy to extend to >2 factor cases, but it is worth noting that in cases where the factors are positively correlated, one would need to take whatever component of the factors which are independent of one another.

5. My intuition is that in cartesian coordinates the R-square between correlated X and Y is actually also the cosine of the angle between the regression lines of X on Y and Y on X. But I can't see an obvious derivation, and I'm too lazy to demonstrate it myself. Sorry!

6. Another intuitive dividend is that this makes it clear why you can by R-squared to move between z-scores of correlated normal variables, which wasn't straightforwardly obvious to me.

7. I'd intuit, but again I can't demonstrate, the case for this becomes stronger with highly skewed interventions where almost all the impact is focused in relatively low probability channels, like averting a very specified existential risk.

Roles are Martial Arts for Agency

A long time ago I thought that Martial Arts simply taught you how to fight – the right way to throw a punch, the best technique for blocking and countering an attack, etc. I thought training consisted of recognizing these attacks and choosing the correct responses more quickly, as well as simply faster/stronger physical execution of same. It was later that I learned that the entire purpose of martial arts is to train your body to react with minimal conscious deliberation, to remove “you” from the equation as much as possible.

The reason is of course that conscious thought is too slow. If you have to think about what you’re doing, you’ve already lost. It’s been said that if you had to think about walking to do it, you’d never make it across the room. Fighting is no different. (It isn’t just fighting either – anything that requires quick reaction suffers when exposed to conscious thought. I used to love Rock Band. One day when playing a particularly difficult guitar solo on expert I nailed 100%… except “I” didn’t do it at all. My eyes saw the notes, my hands executed them, and no where was I involved in the process. It was both exhilarating and creepy, and I basically dropped the game soon after.)

You’ve seen how long it takes a human to learn to walk effortlessly. That's a situation with a single constant force, an unmoving surface, no agents working against you, and minimal emotional agitation. No wonder it takes hundreds of hours, repeating the same basic movements over and over again, to attain even a basic level of martial mastery. To make your body react correctly without any thinking involved. When Neo says “I Know Kung Fu” he isn’t surprised that he now has knowledge he didn’t have before. He’s amazed that his body now reacts in the optimal manner when attacked without his involvement.

All of this is simply focusing on pure reaction time – it doesn’t even take into account the emotional terror of another human seeking to do violence to you. It doesn’t capture the indecision of how to respond, the paralysis of having to choose between outcomes which are all awful and you don’t know which will be worse, and the surge of hormones. The training of your body to respond without your involvement bypasses all of those obstacles as well.

This is the true strength of Martial Arts – eliminating your slow, conscious deliberation and *acting* while there is still time to do so.

Roles are the Martial Arts of Agency.

When one is well-trained in a certain Role, one defaults to certain prescribed actions immediately and confidently. I’ve acted as a guy standing around watching people faint in an overcrowded room, and I’ve acted as the guy telling people to clear the area. The difference was in one I had the role of Corporate Pleb, and the other I had the role of Guy Responsible For This Shit. You know the difference between the guy at the bar who breaks up a fight, and the guy who stands back and watches it happen? The former thinks of himself as the guy who stops fights. They could even be the same guy, on different nights. The role itself creates the actions, and it creates them as an immediate reflex. By the time corporate-me is done thinking “Huh, what’s this? Oh, this looks bad. Someone fainted? Wow, never seen that before. Damn, hope they’re OK. I should call 911.” enforcer-me has already yelled for the room to clear and whipped out a phone.

Roles are the difference between Hufflepuffs gawking when Neville tumbles off his broom (Protected), and Harry screaming “Wingardium Leviosa” (Protector). Draco insulted them afterwards, but it wasn’t a fair insult – they never had the slightest chance to react in time, given the role they were in. Roles are the difference between Minerva ordering Hagrid to stay with the children while she forms troll-hunting parties (Protector), and Harry standing around doing nothing while time slowly ticks away (Protected). Eventually he switched roles. But it took Agency to do so. It took time.

Agency is awesome. Half this site is devoted to becoming better at Agency. But Agency is slow. Roles allow real-time action under stress.

Agency has a place of course. Agency is what causes us to decide that Martial Arts training is important, that has us choose a Martial Art, and then continue to train month after month. Agency is what lets us decide which Roles we want to play, and practice the psychology and execution of those roles. But when the time for action is at hand, Agency is too slow. Ensure that you have trained enough for the next challenge, because it is the training that will see you through it, not your agenty conscious thinking.

As an aside, most major failures I’ve seen recently are when everyone assumed that *someone else* had the role of Guy In Charge If Shit Goes Down. I suggest that, in any gathering of rationalists, they begin the meeting by choosing one person to be Dictator In Extremis should something break. Doesn’t have to be the same person as whoever is leading. Would be best if it was someone comfortable in the role and/or with experience in it. But really there just needs to **be one**. Anyone.

cross-posted from my blog

[LINK] Speed superintelligence?

From Toby Ord:

> Tool assisted speedruns (TAS) are when people take a game and play it frame by frame, effectively providing super reflexes and forethought, where they can spend a day deciding what to do in the next 1/60th of a second if they wish. There are some very extreme examples of this, showing what can be done if you really play a game perfectly. For example, this video shows how to winSuper Mario Bros 3 in 11 minutes. It shows how different optimal play can be from normal play. In particular, on level 8-1, it gains 90 extra lives by a sequence of amazing jumps.
>
> Other TAS runs get more involved and start exploiting subtle glitches in the game. For example, this page talks about speed running NetHack, using a lot of normal tricks, as well as luck manipulation (exploiting the RNG) and exploiting a dangling pointer bug to rewrite parts of memory.

Though there are limits to what AIs could do with sheer speed, it's interesting that great performance can be achieved with speed alone, that this allows different strategies from usual ones, and that it allows the exploitation of otherwise unexploitable glitches and bugs in the setup.

LW client-side comment improvements

All of these things I mentioned in the most recent open thread, but since the first one is directly relevant and the comment where I posted it somewhat hard to come across, I figured I'd make a post too.

Custom Comment Highlights

NOTE FOR FIREFOX USERS: this contained a bug which has been squashed, causing the list of comments not to be automatically populated (depending on your version of Firefox). I suggest reinstalling. Sorry, no automatic updates unless you use the Chrome extension (though with >50% probability there will be no further updates).

You know how the highlight for new comments on Less Wrong threads disappears if you reload the page, making it difficult to find those comments again? Here is a userscript you can install to fix that (provided you're on Firefox or Chrome). Once installed, you can set the date after which comments are highlighted, and easily scroll to new comments. See screenshots. Installation is straightforward (especially for Chrome, since I made an extension as well).

Bonus: works even if you're logged out or don't have an account, though you'll have to set the highlight time manually.

Delay Before Commenting

Another script to add a delay and checkbox reading "In posting this, I am making a good-faith contribution to the collective search for truth." before allowing you to comment. Made in response to a comment by army1987.

Slate Star Codex Comment Highlighter

Edit: You no longer need to install this, since Scott's added it to his blog. Unless you want the little numbers in the title bar.

Yet another script, to make finding recent comments over at Slate Star Codex a lot easier. Also comes in Chrome extension flavor. See screenshots. Not directly relevant to Less Wrong, but there's a lot of overlap in readership, so you may be interested.

Note for LW Admins / Yvain

These would be straightforward to make available to all users (on sufficiently modern browsers), since they're just a bit of Javascript getting injected. If you'd like to, feel free, and message me if I can be of help.

Fighting Biases and Bad Habits like Boggarts

TL;DR: Building humor into your habits for spotting and correcting errors makes the fix more enjoyable, easier to talk about and receive social support, and limits the danger of a contempt spiral.

One of the most reliably bad decisions I've made on a regular basis is the choice to stay awake (well, "awake") and on the internet past the point where I can get work done, or even have much fun. I went through a spell where I even fell asleep on the couch more nights than not, unable to muster the will or judgement to get up and go downstairs to bed.

I could remember (even sometimes in the moment) that this was a bad pattern, but, the more tired I was, the more tempting it was to think that I should just *buckle down* and *apply more willpower* to be more awake and get more out of my computer time. Going to bed was *a* solution, but it was hard for it not to feel (to my sleepy brain and my normal one) like a bit of a cop out.

Only two things helped me really keep this failure mode in check. One was setting a hard bedtime (and beeminding it) as part of my sacrifice for Advent. But the other key tool (which has lasted me long past Advent) is the gif below.

The poor kid struggling to eat his ice cream cone, even in the face of his exhaustion, is hilarious. And not too far off the portrait of me around 2am scrolling through my Feedly.

Thinking about how *stupid* or *ineffective* or *insufficiently strong-willed* I'm being makes it hard for me to do anything that feels like a retreat from my current course of action. I want to master the situation and prove I'm stronger. But catching on to the fact that my current situation (of my own making or not) is ridiculous, makes it easier to laugh, shrug, and move on.

I think the difference is that it's easy for me to feel contemptuous of myself when frustrated, and easy to feel fond when amused.

I've tried to strike the new emotional tone when I'm working on catching and correcting other errors. (e.g "Stupid, you should have known to leave more time to

make the appointment! Planning fallacy!" becomes "Heh, I guess you thought that adding two "trivially short" errands was a closed set, and must remain 'trivially short.' That's a pretty silly error.")

In the first case, noticing and correcting an error feels punitive, since it's quickly followed by a hefty dose of flagellation, but the second comes with a quick laugh and a easier shift to a growth mindset framing. Funny stories about errors are also easier to tell, increasing the chance my friends can help catch me out next time, or that I'll be better at spotting the error just by keeping it fresh in my memory. Not to mention, in order to get the joke, I tend to look for a more specific cause of the error than stupid/lazy/etc.

As far as I can tell, it also helps that *amusement* is a pretty different feeling than the ones that tend to be active when I'm falling into error (frustration, anger, feeling trapped, impatience, etc). So, for a couple of seconds at least, I'm *out* of the rut and now need to actively *return* to it to stay stuck.

In the heat of the moment of anger/akrasia/etc is a bad time to figure out what's funny, but, if you're reflecting on your errors after the fact, in a moment of consolation, it's easier to go back armed with a helpful reframing, ready to cast *Riddikulus*!

Crossposted from my personal blog, Unequally Yoked.

Quantified Risks of Gay Male Sex

If you are a gay male then you've probably worried at one point about sexually transmitted diseases. Indeed men who have sex with men have some of the highest prevalence of many of these diseases. And if you're not a gay male, you've probably still thought about STDs at one point. But how much should you worry? There are many organizations and resources that will tell you to wear a condom, but very few will tell you the relative risks of wearing a condom vs not. I'd like to provide a concise summary of the risks associated with gay male sex and the extent to which these risks can be reduced. (See Mark Manson's guide for a similar resources for heterosexual sex.). I will do so by first giving some information about each disease, including its prevalence among gay men. Most of this data will come from the US, but the US actually has an unusually high prevalence for many diseases. Certainly HIV is much less common in many parts of Europe. I will end with a case study of HIV, which will include an analysis of the probabilities of transmission broken down by the nature of sex act and a discussion of risk reduction techniques.

When dealing with risks associated with sex, there are few relevant parameters. The most common is the prevalence – the proportion of people in the population that have the disease. Since you can only get a disease from someone who has it, the prevalence is arguably the most important statistic. There are two more relevant statistics – the per act infectivity (the chance of contracting the disease after having sex once) and the per partner infectivity (the chance of contracting the disease after having sex with one partner for the duration of the relationship). As it turns out the latter two probabilities are very difficult to calculate. I only obtained those values for for HIV. It is especially difficult to determine per act risks for specific types of sex acts since many MSM engage in a variety of acts with multiple partners. Nevertheless estimates do exist and will explored in detail in the HIV case study section.

HIV

Prevalence: Between 13 - 28%. My guess is about 13%.

The most infamous of the STDs. There is no cure but it can be managed with anti-retroviral therapy. A commonly reported statistic is that 19% of MSM (men who have sex with men) in the US are HIV positive (1). For black MSM, this number was 28% and for white MSM this number was 16%. This is likely an overestimate, however, since the sample used was gay men who frequent bars and clubs. My estimate of 13% comes from CDC's total HIV prevalence in gay men of 590,000 (2) and their data suggesting that MSM comprise 2.9% of men in the US (3).

Gonorrhea

Prevalence: Between 9% and 15% in the US

This disease affects the throat and the genitals but it is treatable with antibiotics. The CDC estimates 15.5% prevalence (4). However, this is likely an overestimate since the sample used was gay men in health clinics. Another sample (in San Francisco health clinics) had a pharyngeal gonorrhea prevalence of 9% (5).

Syphilis

Prevalence: 0.825% in the US

My estimate was calculated in the same manner as my estimate for HIV. I used the CDC's data (6). Syphilis is transmittable by oral and anal sex (7) and causes genital sores that may look harmless at first (8). Syphilis is curable with penicillin however the presence of sores increases the infectivity of HIV.

Herpes (HSV-1 and HSV-2)

Prevalence: HSV-2 - 18.4% (9); HSV-1 - ~75% based on Australian data (10)

This disease is mostly asymptomatic and can be transmitted through oral or anal sex. Sometimes sores will appear and they will usually go away with time. For the same reason as syphilis, herpes can increase the chance of transmitting HIV. The estimate for HSV-1 is probably too high. Snowball sampling was used and most of the men recruited were heavily involved in organizations for gay men and were sexually active in the past 6 months. Also half of them reported unprotected anal sex in the past six months. The HSV-2 sample came from a random sample of US households (11).

Clamydia

Prevalence: Rectal - 0.5% - 2.3% ; Pharyngeal - 3.0 - 10.5% (12)

Like herpes, it is often asymptomatic - perhaps as low as 10% of infected men report symptoms. It is curable with antibiotics.

HPV

Prevalence: 47.2% (13)

This disease is incurable (though a vaccine exists for men and women) but usually asymptomatic. It is capable of causing cancers of the penis, throat and anus. Oddly there are no common tests for HPV in part because there are many strains (over 100) most of which are relatively harmless. Sometimes it goes away on its own (14). The prevalence rate was oddly difficult to find, the number I cited came from a sample of men from Brazil, Mexico and the US.

Case Study of HIV transmission; risks and strategies for reducing risk

IMPORTANT: None of the following figures should be generalized to other diseases. Many of these numbers are not even the same order of magnitude as the numbers for other diseases. For example, HIV is especially difficult to transmit via oral sex, but Herpes can very easily be transmitted.

Unprotected Oral Sex per-act risk (with a positive partner or partner of unknown serostatus):

Non-zero but very small. Best guess .03% without condom (15)

Unprotected Anal sex per-act risk (with positive partner):

Receptive: 0.82% - 1.4% (16) (17)

Insertive Circumcised: 0.11% (18)

Insertive Uncircumcised: 0.62% (18)

Protected Anal sex per-act risk (with positive partner):

Estimates range from 2 times lower to twenty times lower (16) (19) and the risk is highly dependent on the slippage and breakage rate.

Contracting HIV from oral sex is very rare. In one study, 67 men reported performing oral sex on at least one HIV positive partner and none were infected (20). However, transmission is possible (15). Because instances of oral transmission of HIV are so rare, the risk is hard to calculate so should be taken with a grain of salt. The number cited

was obtained from a group of individuals that were either HIV positive or high risk for HIV. The per act-risk with a positive partner is therefore probably somewhat higher.

Note that different HIV positive men have different levels of infectivity hence the wide range of values for per-act probability of transmission. Some men with high viral loads (the amount of HIV in the blood) may have an infectivity of greater than 10% per unprotected anal sex act (17).

Risk reducing strategies

Choosing sex acts that have a lower transmission rate (oral sex, protected insertive anal sex, non-insertive) is one way to reduce risk. Monogamy, testing, antiretroviral therapy, PEP and PrEP are five other ways.

Testing Your partner/ Monogamy

If your partner tests negative then they are very unlikely to have HIV. There is a 0.047% chance of being HIV positive if they tested negative using a blood test and a 0.29% chance of being HIV positive if they tested negative using an oral test. If they did further tests then the chance is even lower. (See the section after the next paragraph for how these numbers were calculated).

So if your partner tests negative, the real danger is not the test giving an incorrect result. The danger is that your partner was exposed to HIV before the test, but his body had not started to make antibodies yet. Since this can take weeks or months, it is possible for your partner who tested negative to still have HIV even if you are both completely monogamous.

———

For tests, the sensitivity - the probability that an HIV positive person will test positive - is 99.68% for blood tests (21), 98.03% with oral tests. The specificity - the probability that an HIV negative person will test negative - is 99.74% for oral tests and 99.91% for blood tests. Hence the probability that a person who tested negative will actually be positive is:

P(Positive | tested negative) = P(Positive)*(1-sensitivity)/(P(Negative)*specificity + P(Positive)*(1-sensitivity)) = 0.047% for blood test, 0.29% for oral test

Where P(Positive) = Prevalence of HIV, I estimated this to be 13%.

However, according to a writer for About.com (22) - a doctor who works with HIV - there are often multiple tests which drive the sensitivity up to 99.997%.

Home Testing

Oraquick is an HIV test that you can purchase online and do yourself at home. It costs $39.99 for one kit. The sensitivity is 93.64%, the specificity is 99.87% (23). The probability that someone who tested negative will actually be HIV positive is 0.94%. - assuming a 13% prevalence for HIV. The same danger mentioned above applies - if the infection occurred recently the test would not detect it.

Anti-Retroviral therapy

Highly active anti-retroviral therapy (HAART), when successful, can reduce the viral load – the amount of HIV in the blood - to low or undetectable levels. Baggaley et. al (17) reports that in heterosexual couples, there have been some models relating viral load to infectivity. She applies these models to MSM and reports that the per-act risk for

unprotected anal sex with a positive partner should be 0.061%. However, she notes that different models produce very different results thus this number should be taken with a grain of salt.

Post-Exposure Prophylaxis (PEP)

A last resort if you think you were exposed to HIV is to undergo post-exposure prophylaxis within 72 hours. Antiretroviral drugs are taken for about a month in the hopes of preventing the HIV from infecting any cells. In one case controlled study some health care workers who were exposed to HIV were given PEP and some were not, (this was not under the control of the experimenters). Workers that contracted HIV were less likely to have been given PEP with an odds ratio of 0.19 (24). I don't know whether PEP is equally effective at mitigating risk from other sources of exposure.

Pre-Exposure Prophylaxis (PrEP)

This is a relatively new risk reduction strategy. Instead of taking anti-retroviral drugs after exposure, you take anti-retroviral drugs every day in order to prevent HIV infection. I could not find a per-act risk, but in a randomized controlled trial, MSM who took PrEP were less likely to become infected with HIV than men who did not (relative reduction - 41%). The average number of sex partners was 18. For men who were more consistent and had a 90% adherence rate, the relative reduction was better - 73%. (25) (26).

1: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5937a2.htm?s_cid=mm5937a2_w

2: http://www.cdc.gov/hiv/statistics/basics/ataglance.html

3: http://www.cdc.gov/nchs/data/ad/ad362.pdf

4: http://www.cdc.gov/std/stats10/msm.htm

5: http://cid.oxfordjournals.org/content/41/1/67.short

6: http://www.cdc.gov/std/syphilis/STDFact-MSM-Syphilis.htm

7: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5341a2.htm

8: http://www.cdc.gov/std/syphilis/stdfact-syphilis.htm

9: http://journals.lww.com/stdjournal/Abstract/2010/06000/Men_Who_Have_Sex_With_Men_in_the_United_States_.13.aspx

10: http://jid.oxfordjournals.org/content/194/5/561.full

11: http://www.nber.org/nhanes/nhanes-III/docs/nchs/manuals/planop.pdf

12: http://www.cdc.gov/std/chlamydia/STDFact-Chlamydia-detailed.htm

13: http://jid.oxfordjournals.org/content/203/1/49.short

14: http://www.cdc.gov/std/hpv/stdfact-hpv-and-men.htm

15: http://journals.lww.com/aidsonline/pages/articleviewer.aspx?year=1998&issue=16000&article=00004&type=fulltext#P80

16: http://aje.oxfordjournals.org/content/150/3/306.short

17: http://ije.oxfordjournals.org/content/early/2010/04/20/ije.dyq057.full

18: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2852627/

19:

http://journals.lww.com/stdjournal/Fulltext/2002/01000/Reducing_the_Risk_of_Sexual_HIV_Transmission_.7.aspx

20:

http://journals.lww.com/aidsonline/Fulltext/2002/11220/Risk_of_HIV_infection_attributable_to_oral_sex.22.aspx

21: http://www.thelancet.com/journals/laninf/article/PIIS1473-3099%2811%2970368-1/abstract

22:

http://aids.about.com/od/hivpreventionquestions/f/How-Often-Do-False-Positive-And-False-Negative-Hiv-Test-Results-Occur.htm

23: http://www.ncbi.nlm.nih.gov/pubmed/18824617

24: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD002835.pub3/abstract

25: http://www.nejm.org/doi/full/10.1056/Nejmoa1011205#t=articleResults

26: http://www.cmaj.ca/content/184/10/1153.short

Six Plausible Meta-Ethical Alternatives

In this post, I list six metaethical possibilities that I think are plausible, along with some arguments or plausible stories about how/why they might be true, where that's not obvious. A lot of people seem fairly certain in their metaethical views, but I'm not and I want to convey my uncertainty as well as some of the reasons for it.

1. Most intelligent beings in the multiverse share similar preferences. This came about because there are facts about what preferences one should have, just like there exist facts about what decision theory one should use or what prior one should have, and species that manage to build intergalactic civilizations (or the equivalent in other universes) tend to discover all of these facts. There are occasional paperclip maximizers that arise, but they are a relatively minor presence or tend to be taken over by more sophisticated minds.
2. Facts about what everyone should value exist, and most intelligent beings have a part of their mind that can discover moral facts and find them motivating, but those parts don't have full control over their actions. These beings eventually build or become rational agents with values that represent compromises between different parts of their minds, so most intelligent beings end up having shared moral values along with idiosyncratic values.
3. There aren't facts about what everyone should value, but there are facts about how to translate non-preferences (e.g., emotions, drives, fuzzy moral intuitions, circular preferences, non-consequentialist values, etc.) into preferences. These facts may include, for example, what is the right way to deal with ontological crises. The existence of such facts seems plausible because if there were facts about what is rational (which seems likely) but no facts about how to *become* rational, that would seem like a strange state of affairs.
4. None of the above facts exist, so the only way to become or build a rational agent is to just think about what preferences you want your future self or your agent to hold, until you make up your mind in some way that depends on your psychology. But at least this process of reflection is convergent at the individual level so each person can reasonably call the preferences that they endorse after reaching reflective equilibrium their morality or real values.
5. None of the above facts exist, and reflecting on what one wants turns out to be a divergent process (e.g., it's highly sensitive to initial conditions, like whether or not you drank a cup of coffee before you started, or to the order in which you happen to encounter philosophical arguments). There are still facts about rationality, so at least agents that are already rational can call their utility functions (or the equivalent of utility functions in whatever decision theory ends up being the right one) their real values.
6. There aren't any normative facts at all, including facts about what is rational. For example, it turns out there is no one decision theory that does better than every other decision theory in every situation, and there is no obvious or widely-agreed-upon way to determine which one "wins" overall.

(Note that for the purposes of this post, I'm concentrating on morality in the axiological sense (what one should value) rather than in the sense of cooperation and compromise. So alternative 1, for example, is not intended to include the possibility that most intelligent beings end up merging their preferences through some kind of grand acausal bargain.)

It may be useful to classify these possibilities using labels from academic philosophy. Here's my attempt: 1. realist + internalist 2. realist + externalist 3. relativist 4. subjectivist 5. moral anti-realist 6. normative anti-realist. (A lot of debates in metaethics concern the meaning of ordinary moral language, for example whether they refer to facts or merely express attitudes. I mostly ignore such debates in the above list, because it's not clear what implications they have for the questions that I care about.)

One question LWers may have is, where does Eliezer's metathics fall into this schema? Eliezer says that there *are* moral facts about what values every intelligence in the multiverse should have, but only humans are likely to discover these facts and be motivated by them. To me, Eliezer's use of language is counterintuitive, and since it seems plausible that there are facts about what everyone should value (or how each person should translate their non-preferences into preferences) that most intelligent beings can discover and be at least somewhat motivated by, I'm reserving the phrase "moral facts" for these. In my language, I think 3 or maybe 4 is probably closest to Eliezer's position.

Hal Finney has just died.

Hal Finney has just died.

http://lists.extropy.org/pipermail/extropy-chat/2014-August/082585.html

http://lesswrong.com/lw/1ab/dying_outside/

I'm very sad to see him go.

Multiple Factor Explanations Should Not Appear One-Sided

In Policy Debates Should Not Appear One-Sided, Eliezer Yudkowsky argues that arguments on questions of fact should be one-sided, whereas arguments on policy questions should not:

> On questions of simple fact (for example, whether Earthly life arose by natural selection) there's a legitimate expectation that the argument should be a one-sided battle; the facts themselves are either one way or another, and the so-called "balance of evidence" should reflect this. Indeed, under the Bayesian definition of evidence, "strong evidence" is just that sort of evidence which we only expect to find on one side of an argument.
>
> But there is no reason for complex actions with many consequences to exhibit this onesidedness property.

The reason for this is primarily that natural selection has caused all sorts of observable phenomena. With a bit of ingenuity, we can infer that natural selection has caused them, and hence they become evidence for natural selection. The evidence for natural selection thus has a common cause, which means that we should expect the argument to be one-sided.

In contrast, even if a certain policy, say lower taxes, is the right one, the rightness of this policy does not cause its evidence (or the arguments for this policy, which is a more natural expression), the way natural selection causes its evidence. Hence there is no common cause of all of the valid arguments of relevance for the rightness of this policy, and hence no reason to expect that all of the valid arguments should support lower taxes. If someone nevertheless believes this, the best explanation of their belief is that they suffer from some cognitive bias such as the affect heuristic.

(In passing, I might mention that I think that the fact that moral debates are not one-sided indicates that moral realism is false, since if moral realism were true, moral facts should provide us with one-sided evidence on moral questions, just like natural selection provides us with one-sided evidence on the question how Earthly life arose. This argument is similar to, but distinct from, Mackie's argument from relativity.)

Now consider another kind of factual issues: multiple factor explanations. These are explanations which refer to a number of factors to explain a certain phenomenon. For instance, in his book *Guns, Germs and Steel*, Jared Diamond explains the fact that agriculture first arose in the Fertile Crescent by reference to no less than eight factors. I'll just list these factors briefly without going into the details of how they contributed to the rise of agriculture. The Fertile Crescent had, according to Diamond (ch. 8):

1. big seeded plants, which were
2. abundant and occurring in large stands whose value was obvious,
3. and which were to a large degree hermaphroditic "selfers".
4. It had a higher percentage of annual plants than other Mediterreanean climate zones
5. It had higher diversity of species than other Mediterreanean climate zones.
6. It has a higher range of elevations than other Mediterrenean climate zones
7. It had a great number of domesticable big mammals.

8. The hunter-gatherer life style was not that appealing in the Fertile Crescent

(Note that all of these factors have to do with geographical, botanical and zoological facts, rather than with facts about the humans themselves. Diamond's goal is to prove that agriculture arose in Eurasia due to geographical luck rather than because Eurasians are biologically superior to other humans.)

Diamond does not mention any mechanism that would make it less likely for agriculture to arise in the Fertile Crescent. Hence the score of pro-agriculture vs anti-agriculture factors in the Fertile Crescent is 8-0. Meanwhile no other area in the world has nearly as many advantages. Diamond does not provide us with a definite list of how other areas of the world fared but no non-Eurasian alternative seem to score better than about 5-3 (he is primarily interested in comparing Eurasia with other parts of the world).

Now suppose that we didn't know anything about the rise of agriculture, but that we knew that there were eight factors which could influence it. Since these factors would not be caused by the fact that agriculture first arose in the Fertile Crescent, the way the evidence for natural selection is caused by the natural selection, there would be no reason to believe that these factors were on average positively probabilistically dependent of each other. Under these conditions, one area having all the advantages and the next best lacking three of them is a highly surprising distribution of advantages. On the other hand, this is precisely the pattern that we would expect given the hypothesis that Diamond suffers from confirmation bias or another related bias. His theory is "too good to be true" and which lends support to the hypothesis that he is biased.

In this particular case, some of the factors Diamond lists presumably are positively dependent on each other. Now suppose that someone argues that all of the factors are in fact strongly positively dependent on each other, so that it is not very surprising that they all co-occur. This only pushes the problem back, however, because now we want an explanation of a) what the common cause of all of these dependencies is (it being very improbable that they all would correlate in the absence of such a common cause) and b) how it could be that this common cause increases the probability of the hypothesis via eight independent mechanisms, and doesn't decrease it via any mechanism. (This argument is complicated and I'd be happy on any input concerning it.)

Single-factor historical explanations are often criticized as being too "simplistic" whereas multiple factor explanations are standardly seen as more nuanced. Many such explanations are, however, one-sided in the way Diamond's explanation is, which indicates bias and dogmatism rather than nuance. (Another salient example I'm presently studying is taken from Steven Pinker's *The Better Angels of Our Nature*. I can provide you with the details on demand.*) We should be much better at detecting this kind of bias, since it for the most part goes unnoticed at present.

Generally, the sort of "too good to be true"-arguments to infer bias discussed here are strongly under-utilized. As our knowledge of the systematic and predictable ways our thought goes wrong increase, it becomes easier to infer bias from the structure or pattern of people's arguments, statements and beliefs. What we need is to explicate clearly, preferably using probability theory or other formal methods, what factors are relevant for deciding whether some pattern of arguments, statements or beliefs most likely is the result of biased thought-processes. I'm presently doing research on this and would be happy to discuss these questions in detail, either publicly or via pm.

***Edit: Pinker's argument.** Pinker's goal is to explain why violence has declined throughout history. He lists the following five factors in the last chapter:

- The Leviathan (the increasing influence of the government)
- Gentle commerce (more trade leads to less violence)
- Feminization
- The expanding (moral) circle
- The escalator of reason

He also lists some "important but inconsistent" factors:

- Weaponry and disarmanent (he claims that there are no strong correlations between weapon developments and numbers of deaths)
- Resource and power (he claims that there is little connection between resource distributions and wars)
- Affluence (tight correlations between affluence and non-violence are hard to find)
- (Fall of) religion (he claims that atheist countries and people aren't systematically less violen

This case is interestingly different from Diamond's. Firstly, it is not entirely clear to what extent these five mechanisms are actually different. It could be argued that "the escalator of reason" is a common cause of the other one's: that this causes us to have better self-control, which brings out the better angels of our nature, which essentially is feminization and the expanding circle, and which leads to better control over the social environment (the Leviathan) which in turn leads to more trade.

Secondly, the expression "inconsistent" suggests that the four latter factors are comprised by different sub-mechanisms that play in different directions. That is most clearly seen regarding weaponry and disarmament. Clearly, more efficient weapons leads to more deaths *when they are being used*. That is an important reason why World War II was so comparatively bloody. But it also leads to a lower chance *of the weapons actually being used*. The terrifying power of nuclear weapons is an important reason why they've only been used twice in wars. Hence we here have two different mechanisms playing in different directions.

I do think that "the escalator of reason" is a fundamental cause behind the other mechanisms. But it also presumably has some effects which increases the level of violence. For one thing, more rational people are more effective at what they do, which means they can kill more people if they want to. (It is just that normally, they don't want to do it as often as irrational people.) (We thus have the same structure that we had regarding weaponry.)

Also, in traditional societies, pro-social behaviour is often underwritten by mythologies which have no basis in fact. When these mythologies were dissolved by reason, many feared that chaous would ensue ("when God is dead, everything is permitted"). This did not happen. But it is hard to deny that such mythologies can lead to less violence, and that therefore their dissolution through reason can lead to more violence.

We shouldn't get too caught up in the details of this particular case, however. What is important is, again, that there is something suspicious with only listing mechanisms that play in the one direction. In this case, it is not even hard to find important mechanisms that play in the other direction. In my view, putting them in the other

scale, as it were, leads to a better understanding of how the history of violence has unfolded. That said, I find DavidAgain's counterarguments below interesting.