Best of LessWrong: February 2012

1. On Saying the Obvious
2. My Algorithm for Beating Procrastination
3. Avoid misinterpreting your emotions
4. Feed the spinoff heuristic!
5. Diseased disciplines: the strange case of the inverted chart
6. Fireplace Delusions [LINK]
7. Hearsay, Double Hearsay, and Bayesian Updates
8. Sticky threads?
9. Get Curious
10. Topics from "Procedural Knowledge Gaps"
11. Quantified Health Prize results announced

Best of LessWrong: February 2012

1. On Saying the Obvious
2. My Algorithm for Beating Procrastination
3. Avoid misinterpreting your emotions
4. Feed the spinoff heuristic!
5. Diseased disciplines: the strange case of the inverted chart
6. Fireplace Delusions [LINK]
7. Hearsay, Double Hearsay, and Bayesian Updates
8. Sticky threads?
9. Get Curious
10. Topics from "Procedural Knowledge Gaps"
11. Quantified Health Prize results announced

On Saying the Obvious

Related to: Generalizing from One Example, Connecting Your Beliefs (a call for help), Beware the Unsurprised

The idea of this article is something I've talked about a couple of times in comments. It seems to require more attention.

As a general rule, what is obvious to some people may not be obvious to others. Is this obvious to you? Maybe it was. Maybe it wasn't, and you thought it was because of hindsight bias.

Imagine a substantive Less Wrong comment. It's insightful, polite, easy to understand, and otherwise good. Ideally, you upvote this comment. Now imagine the same comment, only with "obviously" in front. This shouldn't change much, but it does. This word seems to change the comment in multifarious bad ways that I'd rather not try to list.

Uncharitably, I might reduce this whole phenomenon to an example of a mind projection fallacy. The implicit deduction goes like this: "I found <concept> obvious. Thus, <concept> is inherently obvious." The problem is that obviousness, like probability, is in the mind.

The stigma of "obvious" ideas has another problem in preventing things from being said at all. I don't know how common this is, but I've actually been *afraid* of saying things that I thought were obvious, even though ignoring this fear and just posting has yet to result in a poorly-received comment. (That is, in fact, why I'm writing this.)

Even tautologies, which are always obvious in retrospect, can be hard to spot. How many of us would have explicitly realized the weak anthropic principle without Nick Bostrom's help?

And what about implications of beliefs you already hold? These *should* be obvious, and *sometimes* are, but our brains are notoriously bad at putting two and two together. Luke's example was not realizing that an intelligence explosion was imminent until he read the I.J. Good paragraph. I'm glad he provided that example, as it has saved me the trouble of making one.

This is not (to paraphrase Eliezer) a thunderbolt of insight. I bring it up because I propose a few community norms based on the idea:

- Don't be afraid of saying something because it's "obvious". It's like how your teachers always said there are no stupid questions.
- Don't burden your awesome ideas with "obvious but it needs to be said".
- Don't vote down a comment because it says something "obvious" unless you've thought about it for a while. Also, don't shun "obvious" ideas.
- Don't call an idea obvious as though obviousness were an inherent property of the idea. Framing it as a personally obvious thing can be a more accurate way of saying what you're trying to say, but it's hard to do this without looking arrogant. (I suspect this is actually one of the reasons we implicitly treat obviousness as impersonal.)

I'm not sure if these are good ideas, but I think implementing them would decrease the volume of thoughts we cannot think and things we can't say.

My Algorithm for Beating Procrastination

Part of the sequence: The Science of Winning at Life

After three months of practice, I now use a single algorithm to beat procrastination most of the times I face it.[1] It probably won't work for you quite like it did for me, but it's the best advice on motivation I've got, and it's a major reason I'm known for having the "gets shit done" property. There are reasons to hope that we can eventually break the chain of akrasia; maybe this post is one baby step in the right direction.

How to Beat Procrastination explained our best current general *theory* of procrastination, called "temporal motivation theory" (TMT). As an exercise in practical advice backed by deep theories, *this* post explains the *process* I use to beat procrastination — a process implied by TMT.

As a reminder, here's a rough sketch of how motivation works according to TMT:

$$Motivation = \frac{Expectancy \times Value}{Impulsiveness \times Delay}$$

Or, as Piers Steel summarizes:

> Decrease the certainty or the size of a task's reward — its expectancy or its value — and you are unlikely to pursue its completion with any vigor. Increase the delay for the task's reward and our susceptibility to delay — impulsiveness — and motivation also dips.

Of course, my motivation system is more complex than that. P.J. Eby likens TMT (as a guide for beating procrastination) to the "fuel, air, ignition, and compression" plan for starting your car: it might be true, but a more useful theory would include details and mechanism.

That's a fair criticism. Just as an fMRI captures the "big picture" of brain function at low resolution, TMT captures the big picture of motivation. This big picture helps us see where we need to work at the gears-and-circuits level, so we can become the goal-directed consequentialists we'd *like* to be.

So, I'll share my four-step algorithm below, and tackle the gears-and-circuits level in later posts.

Step 1: Notice I'm procrastinating.

This part's easy. I know I *should* do the task, but I feel averse to doing it, or I just don't feel motivated enough to care. So I put it off, even though my prefrontal cortex keeps telling me I'll be better off if I do it *now*. When this happens, I proceed to step 2.

Step 2: Guess which unattacked part of the equation is causing me the most trouble.

Now I get to play detective. Which part of the equation is causing me trouble, here? Does the task have low value because it's boring or painful or too difficult, or because the reward isn't

that great? Do I doubt that completing the task will pay off? Would I have to wait a long time for my reward if I succeeded? Am I particularly impatient or impulsive, either now or in general? Which part of this problem do I need to attack?

Actually, I lied. *I* like to play *army sniper*. I stare down my telescopic sight at the terms in the equation and interrogate them. "Is it *you*, Delay? Huh, motherfucker? Is it *you*? I've shot you before; don't think I won't do it again!"

But not everyone was raised on violent videogames. *You* may prefer a different role-play.

Anyway, I try to figure out where the main problem is. Here are some of the signs I look for:

When I imagine myself doing the task, do I see myself bored and distracted instead of engaged and interested? Is the task uncomfortable, onerous, or painful? Am I nervous about the task, or afraid of what might happen if I undertake it? Has the task's payoff lost its value to me? Perhaps it never had much value to me in the first place? If my answer to any of these questions is "Yes," I'm probably facing the motivation problem of *low value*.

Do I think I'm likely to succeed at the task? Do I think it's within my capabilities? Do I think I'll actually *get* the reward if I *do* succeed? If my answer to any of these questions is "No," I'm probably facing the problem of *low expectancy*.

How much of the reward only comes after a significant delay, and how long is that delay? If most of the reward comes after a big delay, I'm probably the facing the problem of, you guessed it, *delay*.

Do I feel particularly impatient? Am I easily distracted by other tasks, even ones for which I also face problems of low value, low expectancy, or delay? If so, I'm probably facing the problem of *impulsiveness*.

If the task is low value and low expectancy, and the reward is delayed, I run my expected value calculation again. Am I sure I *should* do the task, after all? Maybe I should drop it or delegate it. If after re-evaluation I *still* think I should do the task, then I move to step 3.

Step 3: Try several methods for attacking that specific problem.

Once I've got a plausible suspect in my sights, I fire away with the most suitable ammo I've got for that problem. Here's a quick review of some techniques described in How to Beat Procrastination:

For attacking the problem of low value: Get into a state of flow, perhaps by gamifying the task. Ensure the task has meaning by connecting it to what you value intrinsically. Get more energy. Use reward and punishment. Focus on what you love, wherever possible.

For attacking the problem of low expectancy: Give yourself a series of small, challenging but achieveable goals so that you get yourself into a "success spiral" and expect to succeed. Consume inspirational material. Surround yourself with others who are succeeding. Mentally contrast where you are now and where you want to be.

For attacking the problem of delay: Decrease the reward's delay if possible. Break the task into smaller chunks so you can get rewards each step of the way.

For attacking the problem of impulsiveness: Use precommitment. Set specific and meaningful goals and subgoal and sub-subgoals. Measure your behavior. Build useful habits.

Each of these skills must be learned and practiced first before you can use them. It took me only a few days to learn the mental habit of "mental contrasting," but I spent *weeks* practicing the skill of getting myself into *success spirals*. I've spent *months* trying various

methods for having more energy, but I can do a lot better than I'm doing now. I'm not very good at goal-setting yet.

Step 4: If I'm still procrastinating, return to step 2.

If I've found some successful techniques for attacking the term in the motivation equation I thought was causing me the most trouble, but I'm still procrastinating, I return to step 2 and begin my assault on *another* term in the equation.

When I first began using this algorithm, though, I usually didn't get that far. By the time I had learned mental contrasting or success spirals or whatever tool made the difference, the task was either complete or abandoned. This algorithm only begins to shine, I suspect, once you've come to some level of mastery on most of the subroutines it employs. Then you can quickly employ them and, if you're still procrastinating, immediately employ others, until your procrastination is beaten.

Personal examples

Let me give you some idea of what it looks like for me to use this algorithm:

Building the large 5×5-unit Ikea "Expedit" bookshelf is boring and repetitive, so I made a game of it. I pounded each wooden peg 4 or 5 times, alternating between these two counts no matter how quickly each peg went into its hole, waiting to see if the girl I was with would notice the pattern. She didn't, so after every 10th peg I gave her a kiss, waiting to see if she'd catch *that* pattern. She didn't, so I started kissing her after every 5th peg.[2] Apparently she thought I was just especially amorous that night.

Sometimes, being an executive director just ain't fun. I need to make lots of decisions with large but uncertain consequences — decisions that some people will love and others will hate. This is not as cozy as the quiet researcher's life to which I had been growing accustomed. In many cases, the task of coming to a decision on something is fraught with anxiety and fear, and I procrastinate. In these cases, I remind myself of how the decision is connected to what I care about. I also purposely stoke my passion for the organization's mission by playing epic world-saving music like "Butterflies and Hurricanes" by Muse: "Change everything you are... your number has been called... you've got to be the best, you've got to change the world... your time is now." Then I re-do my VoI and EV calculations again and I god damned *try*.

While researching How to Beat Procrastination, I hired a German tutor. I planned to apply to philosophy graduate schools, which meant I needed to speak Greek, Latin, French, or German, and German philosophy isn't *quite* as universally bad as the others (e.g. see Thomas Metzinger). But I procrastinated when studying, for my reward was *very* uncertain: would I actually go the route of philosophy grad school, and would my knowledge of German help? My reward was also extremely delayed, likely by several years. In the end, I did the expected value calculation more carefully than before, and concluded that I *shouldn't* keep trying to speak my Rs from my throat. It was the right call: I'm now pretty certain I'll never go to philosophy grad school.

Three times, I've started writing books. But each time, the rewards (appreciation, notoriety, money) were so delayed and uncertain that I gave up. Instead, I broke the books into chunks that I could publish as individual articles.[3] Thus, I received *some* reward (appreciation, growing notoriety) after every article, and had relatively high expectancy for this reward (since my goal was no longer so lofty as to be picked up by a major publisher). Breaking it into chunks also allowed me to focus on writing the pieces for which I had the most passion. Along the way, I used many techniques to boost my energy.

Conclusion

The key is to be *prepared* to conquer procrastination by practicing the necessary sub-skills *first*. Build small skills in the right order. You can't play Philip Glass if you haven't first learned how to play scales, how to work the pedals, how to play arpeggios and ostinatos (*lots* of arpeggios and ostinatos), etc. And you can't beat procrastination if you don't have any ammo ready when you've caught the right causal factor in your sights.

The quest toward becoming a goal-directed consequentialist is long and challenging, much like that of becoming a truth-aiming rationalist. But the rewards are great, and the journey has perks. Remember: true agency is rare but powerful. As Michael Vassar says, "Evidence that people are crazy is evidence that things are easier than you think." Millions of projects fail not because they "can't be done" but because the first 5 people who tried them failed due to boring, pedestrian reasons like procrastination or the planning fallacy. People with just a *bit* more agency than normal — people like Benjamin Franklin and Tim Ferriss — have incredible power.

At the end of *Reasons and Persons*, Derek Parfit notes that non-religious ethics is a young field, and thus we may entertain high hopes for what will be discovered and what is possible. But scientific self-help is even younger. We have only just begun our inquiry into procrastination's causes and cures. We don't yet know what is possible. All we can do is try. If you have something to protect, shut up and do the impossible. Things may not be so impossible as you once thought.

Next post: How to Be Happy

Previous post: How to Beat Procrastination

[1] The main areas where I still usually succumb to procrastination are diet and exercise. Luckily, my metabolism is holding out pretty well so far.

[2] Or, it was *something* like this. I can't remember the exact game I played, now.

[3] My abandoned book *Scientific Self Help* turned into my ongoing blog post sequence The Science of Winning at Life. My abandoned book *Ethics and Superintelligence* was broken into chunks that morphed into Singularity FAQ, The Singularity and Machine Ethics, and many posts from No-Nonsense Metaethics and *Facing the Singularity*. My abandoned book *Friendly AI: The Most Important Problem in the World* was broken into pieces that resulted in Existential Risk and some posts of *Facing the Singularity*.

Avoid misinterpreting your emotions

A couple of weeks ago, I was suffering from insomnia. Eventually my inability to fall asleep turned into frustration, which then led to feelings of self-doubt about my life in general. Soon I was wondering about whether I would ever amount to anything, whether any of my various projects would ever end up bearing fruit, and so forth. As usual, I quickly became convinced that my life prospects were dim, and that I should stop being ambitious and settle for some boring but safe path while I still had the chance.

Then I realized that there was no reason for me to believe in this, and I stopped thinking that way. I still felt frustrated about not being able to sleep, but I didn't feel miserable about my chances in life. To do otherwise would have been to *misinterpret my emotions.*

Let me explain what I mean by that. There are two common stereotypes about the role of emotions. The first says that emotions are something irrational, and should be completely disregarded when making decisions. The second says that emotions are basically always right, and one should follow their emotions above all. Psychological research on emotions suggests that the correct answer lies in between: we have emotions for a reason, and we should follow their advice, but not unthinkingly.

The *Information Principle* says that *emotional feelings provide conscious information from unconscious appraisals of situations*[1]. Your brain is *constantly* appraising the situation you happen to be in. It notes things like a passerby having slightly threatening body language, or conversation with some person being easy and free of misunderstandings. There are countless of such evaluations going on all the time, and you aren't consciously aware of them because you don't need to. Your subconscious mind can handle them just fine on its own. The end result of all those evaluations is packaged into a brief summary, which is the only thing that your conscious mind sees directly. That "executive summary" is what you experience as a particular emotional state. The passerby makes you feel slightly nervous and you avoid her, or your conversational partner feels pleasant to talk with and you begin to like him, even though you don't know why.

To some extent, then, your emotions will guide you to act appropriately in various situations, even when you don't know why you feel the way you do. However, it's important to intepret them correctly. Maybe you meet a new person on a good day and feel good when talking with them. Do you feel good because the person is pleasant to be with, or because the weather is pleasant? In general, emotions are only used as a source of information when their informational value is not called into question[2]. If you know that you are sad because of something that happened in the morning, and still feel sad when talking to your friend later on, you don't assume that something about your friend is making you feel sad.

People also pay more attention to their feelings when they think them relevant for the question at hand. For example, moods have a larger impact when people are making decisions for themselves rather than others, who may experience things differently. But by default, people tend to assume that their feelings and emotions are "about" whatever it is that they're thinking about at that moment. If they're not given a reason to presume that their emotions are caused by something else than the issue at hand, they don't.[2]

So here was my mistake. I had been feeling frustrated about my inability to sleep, and my thoughts had wandered to other subjects, such as my life in general. And then I had automatically assumed that because I was feeling frustrated while thinking about my life, my life wasn't going well, so I should reconsider my plans.

In addition to providing information, moods also affect the *way* we think: research suggests that sad moods make us more analytical. Or as Schwarz[2] summarizes:

> When things go smoothly and we face no hurdles in the pursuit of our goals, we are likely to rely on our pre-existing knowledge structures and routines, which served us well in the past. Moreover, we may be willing to take some risk in exploring novel solutions. Once things go wrong, we abandon reliance on our usual routines and focus on the specifics at hand to determine what went wrong and what can be done about it.

So again: I had been trying to sleep, but failed to do so. My failure at the task triggered feelings of frustration. Frustration is a sign that our current approach isn't working, and we should re-evaluate it. In my situation, the right course of action would probably have been to re-evaluate whether I would be getting any sleep at that moment, and spend some time awake until I'd feel more tired again. But I stayed in bed, so my feelings of frustration persisted, and the impulse to re-evaluate things remained. And when my thoughts wandered to other subjects, it was those subjects that my mind started taking apart to find what was wrong with them. The fact that there *wasn't* actually anything wrong with them didn't matter. Some part of my mind presumed, quite reasonably, that if I was feeling frustrated then there *had* to be something wrong with what I was doing, so if I thought otherwise I had to be mistaken. And this line of reasoning would have been correct, had it not been applied to the wrong problem.

I am slowly learning when I should be taking my negative moods into account, and when I shouldn't. I've noticed that on days when I haven't had enough sleep, I also feel skeptical about what I'm doing with my life. When I'm more rested and in a neutral mood, those doubts seem overblown. So I try to discount such doubts when they seem to be caused by mere physical fatigue. On the other hand, some negative feelings are such that I've generally come to regret overriding them. Sometimes I've gotten a bad vibe about a person, and when I've decided to trust them anyway, I've afterwards realized that I shouldn't have.

Positive emotions, too, can be correct or mistaken. I have a tendency to get quite excited about new projects, and be much more certain of their value than I should be. At such times, I try to make sure that I'm not rushing ahead with the project and making commitments that I shouldn't.

Thinking in such a way is an example of taking the outside view. When someone takes the inside view to a problem, such as the task of predicting how long something will take, they *focus on the case at hand, consider the plan and the obstacles to its completion, construct scenarios of future progress, and extrapolate current trends*[3]. On the other hand, the outside view essentially ignores the details of the case at hand, and *involves no attempt at detailed forecasting of the future history of the project. Instead, it focuses on the statistics of a class of cases chosen to be similar in relevant respects to the present one*[3]. For instance, when considering how long it will take you to write an essay, the inside view might respond by looking at how well you've done so far, and how long it would take if you kept up the pace. The outside view would simply look at previous occasions when you've had to write an essay, and

ask how long it took on those occasions. If on several previous occasions you've thought that you'll get the essay written in no time, but then always finished just before the deadline, then it's most likely that you'll again finish right before the deadline.

It's generally beneficial to take the outside view on your emotions as well. In a strongly emotional state, you cannot rely merely on the inside view, because a large part of your reasoning process is working on the basis of assumptions which may not be correct. Instead, you should ask questions like: On previous occasions when you've been in a similar situation and felt similarly, has the advice from your emotions been reasonable? What's their historical accuracy in these circumstances? Is your emotional state being influenced by something that has nothing to do with the issue at hand? If you were a neutral observer looking at the situation from the outside, would you think that the emotional judgement was a reasonable one, or that you were just being silly?

For a long time, I thought that if I was feeling miserable and it was making me think negative thoughts, I only had two options. **A**, I could get rid of the negative thoughts by distracting myself or finding something that would cheer me up and get me out of that mood. Or **B**, I would fail to get out of the mood, and thus keep thinking negative thoughts. For whatever reason, I never realized that I also had option **C**: keep feeling miserable, but stop thinking negative thoughts. Depending on exactly how strong your emotion is, you might not always be capable of getting rid of the thoughts, but at least you can realize that they're not true.

So that's what I did. I thought, "I'm feeling miserable because I can't sleep and I'm frustrated, but that has nothing to do with whether my projects and ambitions will be successful or not. My current emotions convey no information about that topic. So it's pointless to doubt myself because of these emotions." (Not in so many words, but that was the general idea.) So I stopped thinking those thoughts. And while I still felt generally miserable, the thoughts stopped making me feel even worse.

Possibly the most vivid example of taking the outside view that I've seen comes from Ferrett Steinmetz:

> I was suicidally down yesterday for no reason except brain chemistry, waking up with the belief that everyone I knew would be much better off if I killed myself. And I did my usual ration-checks to see if what depression was saying was correct – because, like bullies, occasionally the cruel will tell you what the kind will not. So I looked at the evidence.
>
> What the evidence told me was that as a polyamorous man, I had several women who loved me deeply, women who had the choice of other partners and yet still cared about me enough to send me texts and emails, and this should be evidence that I was not a worthless human being. At which point my depression started in on me: *See? All these women who love you, and you just write them off. That's how selfish you are, ignoring the adoration of these women. You're such a self-centered asshole, you should kill yourself.*
>
> Fortunately, I knew my old adversary well enough to understand where it was leading me. I stepped away from the self-destructive sequence my depression was trying to guide me down, recognizing that when I'm in this mood every path goes straight to off-yourself-ville, and understood that the facts would have to be enough.

> Depression is a bully in that it’s fundamentally out to destroy you. You can’t quite get away from him, like any good bully; the best you can do is come to an understanding that this is unpleasant, but it’s nothing you should take too personally. And hope, one day, that you’ll become strong enough to walk away.

Major depression is the extreme case. If your depression is serious enough, your brain is broken. The mechanisms which would usually kick in when you were doing something wrong will be engaged even when you're doing *nothing* wrong, and they will be in overdrive, taking apart *everything* in your life in order to find ways by which you are screwing up.

But you don't have to believe them. You can realize that the thoughts that pop up in your mind aren't based on reality, and that you don't have to act on the basis of them. It won't stop you from feeling miserable, but it might stop you from feeling even *worse.*

Edited to add: Of course, it's also possible to use this view for self-deception. Maybe we're deceiving ourselves about how our lives are going, and that self-deception will persist if we try to examine it while in a neutral emotional state. Perhaps it is only when we fail badly enough to get a strong negative emotion that the barriers of self-deception break, and we will be mistaken to dismiss our thoughts in those states because they don't seem reasonable in other emotional states. When you use this technique, be careful to make sure that you are actually genuinely curious about what your emotions are telling you. Don't *just* come up with excuses for ignoring them, ask whether you should ignore them *or* listen to them.

References

[1]: Clore, G.L. & Gasper, K., & Garvin, E. (2001). Affect as information. In J.P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 121–144). Mahwah, NJ: Erlbaum.

[2]: Schwarz, N. (2010) Feelings as information. In Van Lange, P. & Kruglanski, A. & Higgins, E.T. (Eds.), *Handbook of theories of social psychology*, Sange.

[3]: Kahneman, D. & Lovallo, D. (1993) Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking. Management Science, vol. 39, no. 1.

Feed the spinoff heuristic!

Follow-up to:

Parapsychology: the control group for science

Some Heuristics for Evaluating the Soundness of the Academic Mainstream in Unfamiliar Fields

Recent renewed discussions of the parapsychology literature and Daryl Bem's recent precognition article brought to mind the "market test" of claims of precognition. Bem tells us that random undergraduate students were able to predict with 53% accuracy where an erotic image would appear in the future. If this effect was actually real, I would rerun the experiment before corporate earnings announcements, central bank interest rate changes, etc, and change the images based on the reaction of stocks and bonds to the announcements. In other words, I could easily convert "porn precognition" into "hedge fund trillionaire precognition."

If I was initially lacking in the capital to do trades, I could publish my predictions online using public key cryptography and amass an impressive track record before recruiting investors. If anti-psi prejudice was a problem, no one need know how I was making my predictions. Similar setups could exploit other effects claimed in the parapsychology literature (e.g. the remote viewing of the Scientologist-founded Stargate Project of the U.S. federal government). Those who assign a lot of credence to psi may want to actually try this, but for me this is an invitation to use parapsychology as control group for science, and to ponder a general heuristic for crudely estimating the soundness of academic fields for outsiders.

One reason we trust that physicists and chemists have some understanding of their subjects is that they produce valuable technological spinoffs with concrete and measurable economic benefit. In practice, I often make use of the spinoff heuristic: If an unfamiliar field has the sort of knowledge it claims, what commercial spinoffs and concrete results ought it to be producing? Do such spinoffs exist? What are the explanations for their absence?

For psychology, I might cite systematic desensitization of specific phobias such as fear of spiders, cognitive-behavioral therapy, and military use of IQ tests (with large measurable changes in accident rates, training costs, etc). In financial economics, I would raise the hundreds of billions of dollars invested in index funds, founded in response to academic research, and their outperformance relative to managed funds. Auction theory powers tens of billions of dollars of wireless spectrum auctions, not to mention evil dollar-auction sites.

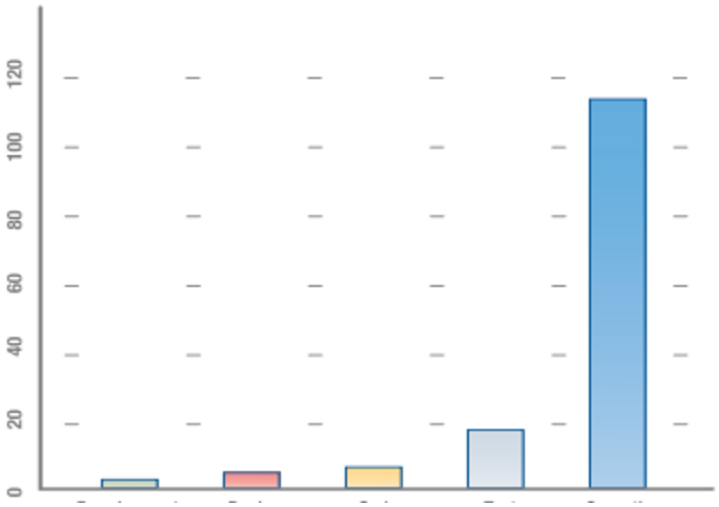
This seems like a great task for crowdsourcing: the cloud of LessWrongers has broad knowledge, and sorting real science from cargo cult science is core to being Less Wrong. So I ask you, Less Wrongers, for your examples of practical spinoffs (or suspicious absences thereof) of sometimes-denigrated fields in the comments. Macroeconomics, personality psychology, physical anthropology, education research, gene-association studies, nutrition research, wherever you have knowledge to share.

ETA: This academic claims to be trying to use the Bem methods to predict roulette wheels, and to have passed statistical significance tests on his first runs. Such claims

have been made for casinos in the past, but always trailed away in failures to replicate, repeat, or make actual money. I expect the same to happen here.

Diseased disciplines: the strange case of the inverted chart

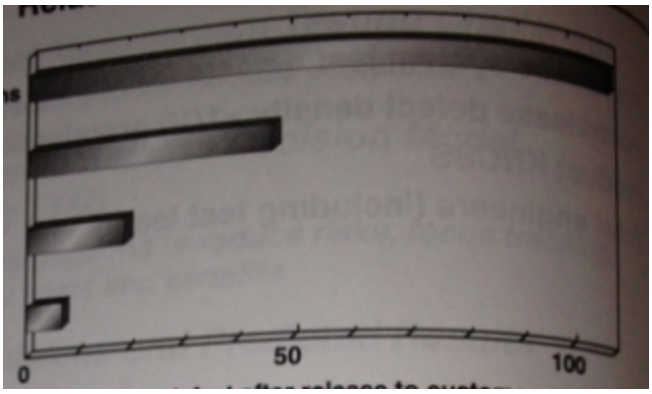
Imagine the following situation: you have come across numerous references to a paper purporting to show that the chances of successfully treating a disease **contracted at age 10** are substantially lower if the disease is **detected later**: somewhat lower at age 20 to very poor at age 50. Every author draws more or less the same bar chart to depict this situation: the picture below, showing **rising** mortality from left to right.



You search for the original paper, which proves a long quest: the conference publisher have lost some of their archives in several moves, several people citing the paper turn out to no longer have a copy, etc. You finally locate a copy of the paper (let's call it G99) thanks to a helpful friend with great scholarly connections.

And you find out some interesting things.

The most striking is what the author's original chart depicts: the chances of successfully treating the disease **detected at age 50** become substantially lower as a function of **age when it was contracted**; mortality is highest if the disease was contracted at age 10 and lowest if contracted at age 40. The chart showing this is the picture below, showing **decreasing** mortality from top to bottom, for the same ages on the vertical axis.



Not only is the representation topsy-turvy; the two diagrams can't be about the same thing, since what is constant in the first (age disease detected) is variable in the other, and what is variable in the first (age disease contracted) is constant in the other.

Now, as you research the issue a little more, you find out that authors prior to G99 have often used the first diagram to report their findings; reportedly, several different studies on different populations (dating back to the eighties) have yielded similar results.

But when citing G99, nobody reproduces the *actual* diagram in G99, they all reproduce the older diagram (or some variant of it).

You are tempted to conclude that the authors citing G99 are citing "from memory"; they are aware of the earlier research, they have a vague recollection that G99 contains results that are not totally at odds with the earlier research. Same difference, they reason, G99 is **one more confirmation** of the earlier research, which is adequately summarized by the standard diagram.

And then you come across a paper by the same author, but from 10 years earlier. Let's call it G89. There is a strong presumption that the study in G99 is the same that is described in G89, for the following reasons: a) the researcher who wrote G99 was by then already retired from the institution where they obtained their results; b) the G99 "paper" isn't in fact a paper, it's a PowerPoint summarizing previous results obtained by the author.

And in G89, you read the following: "This study **didn't accurately record** the mortality rates at various ages after contracting the disease, so we will use average rates summarized from several other studies."

So basically everyone who has been citing G99 has been building castles on sand.

Suppose that, far from some exotic disease affecting a few individuals each year, the disease in question was one of the world's major killers (say, tuberculosis, the world's leader in infectious disease mortality), and the reason why everyone is citing either G99 or some of the earlier research is to lend support to the standard strategies for fighting the disease.

When you look at the earlier research, you find nothing to allay your worries: the earlier studies are described only summarily, in broad overview papers or secondary sources; the numbers don't seem to match up, and so on. In effect you are discovering, about thirty years later, that what was taken for granted as a major

finding on one of the principal topics of the discipline in fact has "sloppy academic practice" written all over it.

If this story was true, and this was medicine we were talking about, what would you expect (or at least hope for, if you haven't become too cynical), should this story come to light? In a well-functioning discipline, a wave of retractations, public apologies, general embarrassment and a major re-evaluation of public health policies concerning this disease would follow.

The story is substantially true, but the field isn't medicine: it is software engineering.

I have transposed the story to medicine, temporarily, as an act of benign deception, to which I now confess. My intention was to bring out the structure of this story, and if, while thinking it was about health, you felt outraged at this miscarriage of academic process, you should still feel outraged upon learning that it is in fact about software.

The "disease" isn't some exotic oddity, but the software equivalent of tuberculosis - the cost of fixing defects (a.k.a. bugs).

The original claim was that "defects introduced in early phases cost more to fix the later they are detected". The misquoted chart says this instead: "defects detected in the operations phase (once software is in the field) cost more to fix the earlier they were introduced".

Any result concerning the "disease" of software bugs counts as a major result, because it affects very large fractions of the population, and accounts for a major fraction of the total "morbidity" (i.e. lack of quality, project failure) in the population (of software programs).

The earlier article by the same author contained the following confession: "This study didn't accurately record the engineering times to fix the defects, so we will use average times summarized from several other studies to weight the defect origins".

Not only is this one major result suspect, but the same pattern of "citogenesis" turns up investigating several other important claims.

Software engineering is a diseased discipline.

The publication I've labeled "G99" is generally cited as: Robert B. Grady, An Economic Release Decision Model: Insights into Software Project Management, in proceedings of Applications of Software Measurement (1999). The second diagram is from a photograph of a hard copy of the proceedings.

Here is one typical publication citing Grady 1999, from which the first diagram is extracted. You can find many more via a Google search. The "this study didn't accurately record" quote is discussed here, and can be found in "Dissecting Software Failures" by Grady, in the April 1989 issue of the "Hewlett Packard Journal"; you can still find one copy of the original source on the Web, as of early 2013, but link rot is threatening it with extinction.

A more extensive analysis of the "defect cost increase" claim is available in my book-in-progress, "The Leprechauns of Software Engineering".

Here is how the axes were originally labeled; first diagram:

- vertical: "Relative Cost to Correct a Defect"
- horizontal: "Development Phase" (values "Requirements", "Design", "Code", "Test", "Operation" from left to right)
- figure label: "Relative cost to correct a requirement defect depending on when it is discovered"

Second diagram:

- vertical: "Activity When Defect was Created" (values "Specifications", "Design", "Code", "Test" from top to bottom)
- horizontal: "Relative cost to fix a defect after release to customers compared to the cost of fixing it shortly after it was created"
- figure label: "Relative Costs to Fix Defects"

Fireplace Delusions [LINK]

Sam Harris, in his recent article called *The Fireplace Delusion*, tries to make you feel what it's like to react to a cached belief being irreparably destroyed. Just incase you forgot what your apostasy (if you had one, of course) was like in its early stages.

What are some of the Fireplace Delusions you've come across in your days?

EDIT: WOODSMOKE HEALTH EFFECTS

Hearsay, Double Hearsay, and Bayesian Updates

Application of: How Much Evidence Does It Take?

(trigger warning: some description of domestic violence)

Summary: I discuss the strengths and weaknesses of one way that the American legal system tries to assess and cope with the unreliability of certain kinds of evidence. After explaining the relevant rules with references to a few recent famous cases and a non-notable case that I'm working on now, I briefly consider whether this part of the evidence code is above or below the sanity waterline, and suggest an incremental improvement.

Recently, I got to the point in my legal career where people are trusting me to write evidentiary briefs, i.e., to argue in front of a judge about what kinds of evidence are reliable enough to be safely presented to a jury. There is an odd division of epistemological labor in the American court system: judges are thought [page 90] to be better than juries at resisting passionate or manipulative oratory, and juries are thought to be better than judges at resisting bribery and (pre-existing) personal hatred. As a result, potentially inflammatory or unreliable evidence is presented first to a judge, who (much like one of Eliezer's Confessors) is supposed to sift the exhibit to see if normal people can handle it without losing their tenuous grip on sanity. If and only if the evidence seems safe for ordinary human consumption, the judge will allow the lawyers to argue about that evidence in front of the jury. Otherwise, the evidence sits in a cardboard box in an unheated warehouse, safely away from the eyes of the jury, until it's time for an appeal.

The Hearsay Rule

By way of a concrete example, one famous recent case featured a recorded 911 call made by a domestic violence victim to the emergency phone operator. The operator asked questions about the location and identity of the person who was accused of beating the caller. The caller answered the questions on tape, explicitly identifying her abuser as Mr. Adrian Martell Davis, and the answers were used first to find and arrest the suspect, and ultimately to convict him. The victim was apparently too intimidated to testify in open court, and so her recorded statement as to the name of her abuser was absolutely necessary to support a conviction -- no recording, no conviction. Under the 400-year-old hearsay rule, recorded testimony typically is **not** allowed to be presented to a jury -- courts are concerned that the person giving the recorded statement might be pressured by the police in ways that wouldn't show up on tape, and that allowing a witness to testify without showing up in court unfairly deprives the defendant of a chance to (a) cross-examine the witness, and (b) have the jury see any facial tics, body language, etc. that undercut the witness's credibility. In the 911 case, though, the Court faced a straight choice between finding an exception to the hearsay rule and letting an apparent abuser go free.

In making this choice, the US Supreme Court managed to ignore a variety of emotionally salient but epistemologically irrelevant distractions, such as the seriousness of the crime, the relative helplessness of the victim, and the respectability

of the 911 operator. Instead, the Court focused on the purpose for which the 911 statements were obtained. If the statements were obtained to help gather information needed to safely resolve an ongoing emergency, they could be used at trial. If the statements, however, were obtained to gather information about a past event, they could *not* be used at trial.

The theory supporting this distinction seems to have been that the right to cross-examine and the right to have the jury see body language are fungible elements of a more general reliability test. A stranger's assertion, without more, could be true or could be false. It doesn't count as very much evidence. To turn an assertion into enough evidence to convict someone beyond a reasonable doubt, you need to show that the assertion comes with "indicia of reliability." Two of these indicia are cross-examination and body language -- if a story checks out despite a vigorous unfriendly interview and the peer pressure of having to tell the story while physically in the room with other people from your community, then that's pretty good evidence. But you might have reasons to believe a story even if you don't get cross-examination or body language. In the case of the 911 call, one might think that the caller had a strong motive to tell the truth, because if she didn't, then the police would go looking for the wrong guy, and her abuser would come find her and continue hurting her. Similarly, one might think that the operators had a strong motive to ask fair, non-leading questions, because of they didn't get the right answer, then the police might show up in the wrong neighborhood or with the wrong expectations, and there could be an unnecessary firefight. Finally, one could argue that a recorded statement made as events were unfolding is inherently more reliable (in some ways) than a narrative given months or years after the event; human memory gets corrupted faster than 8-track tapes.

Some combination of these factors convinced the Court to admit the evidence. Other, very similar cases have been decided differently. Whether they got that particular decision right or wrong, though, the framework of "indicia of reliability" is hard-coded into American evidence law, especially for civil cases. If you want to present evidence to a jury based on a statement that was made outside of court, you have to give at least one reason why the statement is nevertheless reliable.

Double and Triple Hearsay

Here's where things really get interesting: if your out-of-court statement quotes another out-of-court statement, the evidence is called "double hearsay," and you need to independently verify each statement. If any link in the chain breaks, the whole document gets excluded. For example, in the case I'm working on now, the defendants want to show the jury a report filled out by California's Occupational Health and Safety Administration ("OSHA"). The OSHA report is based almost entirely on an accident report form filled out by a private corporation. That report form, in turn, is based almost entirely on an informal interview of the only eyewitness to an accident. So the defendants can use the OSHA report if and only if the OSHA report, the accident report, and the informal interview are all reliable. Use $A \leftrightarrow (A \wedge B \wedge C)$ are reliable.

To try to qualify the OSHA report, the defendants are arguing that the OSHA report is reliable under the public record exception to the hearsay rule, meaning that the public officials who prepared it had a stronger interest in accurately reporting public information than they did in the outcome of the accident victim's private case. To get the accident report form in, the defendants are arguing that it is reliable under the business record exception to the hearsay rule, meaning that the corporate officials

who prepared it had a stronger interest in making sure their company had access to accurate information about safety risks than they did in the outcome of any one customer's lawsuit. As for the informal interview...well, I honestly have no idea how they plan to justify its reliability. But, then again, I'm biased. My professional interest lies in making sure that the whole string of unhelpful quotations stays in a cardboard box in a dank garage, far away from any juries.

Do the Rules Work?

So far, I've been pleasantly surprised at how well the American legal system handles some of these challenges. The fact that we have a two-tiered system of evaluating evidence at all is a cut above average -- imagine, e.g., the doctor who examines you taking notes on your condition, filtering out any subjective comments you make about how you're sure it's just a cold, and reporting only your objective symptoms to a second doctor, who then renders a diagnosis. Or imagine a team of business consultants who interview a Fortune 500 company's leadership team, and then pass their written notes back to a team at HQ (who has never met the executives) so that HQ can catch any obvious mistakes in reasoning before sending out recommendations. We know, intellectually, that meeting people tends to make us friendlier toward them and more likely to adopt their point of view even if we encounter no Bayesian evidence that increases the plausibility of their opinions, but our institutions rarely take steps to guard against that bias.

I think my biggest criticism of the American evidence code is that it doesn't account for uncertainty in the model. For instance, if I read the headline on a piece of science journalism saying that (e.g.) coffee consumption reduces the risk of prostate cancer, or that receiving spankings in childhood is negatively correlated with conscientiousness as an adult, there are least six layers of 'hearsay' -- I might have misunderstood the headline, the headline might have mis-summarized the article, the article might have misquoted the scientist, the scientist might have misinterpreted the recorded data, the recorded data might not faithfully reflect what actually happened during the experiment, and the experiment might not faithfully replicate the real-world conditions that interest us.

Even if I can articulate plausible reasons why each step in the transmission of information was "reliable," I should be very skeptical that my *model* of the transmission is accurate. I only have to be wrong about one of the six steps for my estimate of the information's plausibility to be untrustworthy. If the information would only provide a few decibels of evidence even if it were perfectly reliable, then trying to calculate how many points a semi-reliable piece of evidence is worth can fail because of a low signal-to-noise ratio. E.g., suppose I learn that neither the suspect nor the actual criminal were redheads - I might be absolutely certain of this new piece of information, but that's still nowhere near enough evidence to support a conviction. If instead I learn that there is probably something like a 60% chance that neither the suspect nor the criminal had red hair, that datum really doesn't tell me anything at all -- the info shouldn't shift my prior enough for my prior to be noticeably different.

Although courts are allowed to consider the extent to which an unduly long chain of inferences makes evidence less "trustworthy," I think that on balance decisions would be more accurate if there were a firm limit -- say, three layers -- beyond which evidence was simply inadmissible as a matter of law. If A says that B says that C says that D shot someone, then no matter how reliable we think A, B, and C are, we should probably keep that evidence away from the jury unless we can haul at least one of B, C, or D into court to answer cross-examination.

Sticky threads?

It annoys me that there's no way to sticky a thread in the discussion section.

Therefore, I propose creating an **LW wiki page called "Stickies"**, where sticky-worthy threads would be linked to. Would that be acceptable?

These are the threads I'm planning to add:

- the current Welcome to LW thread
- the current MoR Discussion thread
- the current Rationality Quotes thread (OK, they're posted in Main, but still...)
- the current Open Thread
- the current Media Thread
- the current 'What are you working on?' thread

ETA: Following a tip-off by peaigr, I re-purposed the Special Threads wiki page for this. (The stickies are in the 'periodic threads' section.) Now if there were a way to make this page more conspicuous...

Meanwhile, dbaupp has submitted a feature request.

Get Curious

> Being levels above in [rationality] means doing rationalist practice 101 much better than others [just like] being a few levels above in fighting means executing a basic front-kick much better than others.
>
> - lessdazed

> I fear not the man who has practiced 10,000 kicks once, but I fear the man who has practiced one kick 10,000 times.
>
> - Bruce Lee

Recently, when Eliezer wanted to explain why he thought Anna Salamon was among the best rationalists he knew, he picked out one feature of Anna's behavior in particular:

> I see you start to answer a question, and then you stop, and I see you *get curious*.

For me, the ability to reliably *get curious* is the basic front-kick of epistemic rationality. The best rationalists I know are not necessarily those who know the finer points of cognitive psychology, Bayesian statistics, and Solomonoff Induction. The best rationalists I know are those who can reliably *get curious*.

Once, I explained the Cognitive Reflection Test to Riley Crane by saying it was made of questions that tempt your intuitions to quickly give a *wrong* answer. For example:

> A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

If you haven't seen this question before and you're like most people, your brain screams "10 cents!" But elementary algebra shows that can't be right. The correct answer is *5* cents. To get the right answer, I explained, you need to *interrupt* your intuitive judgment and think "No! *Algebra*."

A *lot* of rationalist practice is like that. Whether thinking about physics or sociology or relationships, you need to catch your intuitive judgment and think "No! *Curiosity.*"

Most of us know how to do algebra. How does one "do" curiosity?

Below, I propose a process for how to "get curious." I think we are only *just beginning* to learn how to create curious people, so please don't take this method as Science or Gospel but instead as an attempt to Just Try It.

As with my algorithm for beating procrastination, you'll want to practice each step of the process *in advance* so that when you want to get curious, you're well-practiced on each step already. With enough practice, these steps may even become *habits*.

Step 1: Feel that you don't already know the answer.

If you have beliefs about the matter already, push the "reset" button and erase that part of your map. You must *feel* that you don't already know the answer.

Exercise 1.1: Import the feeling of uncertainty.

1. Think of a question you clearly *don't* know the answer to. When will AI be created? Is my current diet limiting my cognitive abilities? Is it harder to become the Prime Minister of Britain or the President of France?
2. Close your eyes and pay attention to how that blank spot on your map *feels*. (To me, it feels like I can see a silhouette of someone in the darkness ahead, but I wouldn't take bets on who it is, and I expect to be surprised by their identity when I get close enough to see them.)
3. Hang on to that feeling or image of uncertainty and think about the thing you're trying to get curious about. If your old certainty creeps back, switch to thinking about who composed the Voynich manuscript again, then import that feeling of uncertainty into the thing you're trying to get curious about, again.

Exercise 1.2: Consider all the things you've been confident but wrong about.

1. Think of things you once believed but were wrong about. The more similar those beliefs are to the beliefs you're now considering, the better.
2. Meditate on the frequency of your errors, and on the depths of your biases (if you know enough cognitive psychology).

Step 2: Want to know the answer.

Now, you must *want* to fill in this blank part of your map.

You mustn't wish it to remain blank due to apathy or fear. Don't *avoid* getting the answer because you might learn you should eat less pizza and more half-sticks of butter. Curiosity seeks to annihilate itself.

You also mustn't let your desire that your inquiry have a certain answer block you from discovering how the world actually is. You must want your map to resemble the territory, *whatever* the territory looks like. This enables you to change things more effectively than if you falsely believed that the world was already the way you want it to be.

Exercise 2.1: Visualize the consequences of being wrong.

1. Generate hypotheses about the ways the world may be. Maybe you should eat less gluten and more vegetables? Maybe a high-protein diet plus some nootropics would boost your IQ 5 points? Maybe your diet is fairly optimal for cognitive function already?
2. Next, visualize the consequences of being wrong, including the consequences of remaining ignorant. Visualize the consequences of performing 10 IQ points below your potential because you were too lazy to investigate, or because you were strongly motivated to justify your preference for a particular theory of nutrition. Visualize the consequences of screwing up your neurology by taking nootropics you feel excited about but that often cause harm to people with cognitive architectures similar to your own.

Exercise 2.2: Make plans for different worlds.

1. Generate hypotheses about the way the world could be — different worlds you might be living in. Maybe you live in a world where you'd improve your cognitive

function by taking nootropics, or maybe you live in a world where the nootropics would harm you.
2. Make plans for what you'll do if you happen to live in World #1, what you'll do if you happen to live in World #2, etc. (For unpleasant possible worlds, this also gives you an opportunity to leave a line of retreat for yourself.)
3. Notice that these plans are *different*. This should produce in you some curiosity about which world you *actually* live in, so that you can make plans appropriate for the world you *do* live in rather than for one of the worlds you *don't* live in.

Exercise 2.3: Recite the Litany of Tarski.

The Litany of Tarski can be adapted to any question. If you're considering whether the sky is blue, the Litany of Tarski is:

If the sky is blue
I desire to believe the sky is blue.
If the sky is not blue,
I desire not to believe the sky is blue.

Exercise 2.4: Recite the Litany of Gendlin.

The Litany of Gendlin reminds us:

What is true is already so.
Owning up to it doesn't make it worse.
Not being open about it
doesn't make it go away.
And because it's true,
it is what is there to be interacted with.
Anything untrue isn't there to be lived.
People can stand what is true,
for they are already enduring it.

Step 3: Sprint headlong into reality.

If you've made yourself uncertain and then curious, you're now in a position to use argument, empiricism, and scholarship to sprint headlong into reality. This part probably requires some domain-relevant knowledge and an understanding of probability theory and value of information calculations. What tests could answer your question quickly? How can you perform those tests? If the answer can be looked up in a book, which book?

These are important questions, but I think the first two steps of getting curious are more important. If someone can master steps 1 and 2, they'll be so driven by curiosity that they'll eventually figure out how to do step 3 for many scenarios. In contrast, most people who are equipped to do step 3 pretty well still get the wrong answers because they can't reliably execute steps 1 and 2.

Conclusion: Curiosity in Action

A burning itch to know is higher than a solemn vow to pursue truth. If you think it is your *duty* to doubt your own beliefs and criticize your own arguments, then you may do this for a while and conclude that you have done your duty and you're a Good

Rationalist. Then you can feel satisfied and virtuous and move along without being genuinely curious.

In contrast,

> if you can find within yourself the slightest shred of true uncertainty, then guard it like a forester nursing a campfire. If you can make it blaze up into a flame of curiosity, it will make you light and eager, and give purpose to your questioning and direction to your skills.

My recommendation? Practice the front-kick of epistemic rationality *every day*. For months. Train your ape-brain to *get curious*.

Rationality is not magic. For many people, it can be learned and trained.

Topics from "Procedural Knowledge Gaps"

About a year ago, we had a major discussion about procedural knowledge gaps. Here's what was covered....

How to tell whether food is fresh

pjeby claims that eating raw chicken is safe because the gag reflex identifies it fast

How to buy investments

Memorizing the alphabet and other arbitrary lists

Comparison of various how-to sites

General discussion of making things (including Less Wrong) easier to use

How jump start a stalled car.

How to use the Yellow Pages.

Cheap and easy healthy food.

Questions about preparing a simple soup.

Exercise.

Starting relationships, especially for heterosexual men.

Cooking in general.

Browning meat.

How to become bisexual.

How to transfer money from one electronic account to another.

How to buy a used car.

Interacting with police. (Don't talk to US police! The rules are different in the UK.)

How to speak clearly, slowly, etc..

How to fold a fitted sheet.

How to make a will.

How to order at a bar. Also, some cookbook recommendations.

Tipping in the US.

Tipping in the UK and France.

Spacial orientation.

Personal hygiene-- washing, soap, shampoo.

Haircuts for men.

Haircuts for women.

Growing and maintaining long hair.

Putting a cover on a duvet.

Telling the difference between flirting and friendliness.

Choosing shirts that fit.

Left vs. Right (which hand, not political).

Shaving one's face.

How to end conversations politely.

How to make people laugh.

Mailing large objects in the US.

How to format comments at Less Wrong.

How to declutter.

E readers.

Touch typing.

Dvorak, etc.

How often to see a doctor.

Remembering to be polite.

Tying shoes.

Home maintenance.

What might melt in a dishwasher.

Kitchen knives.

Sorting laundry.

Does cranking up the thermostat heat the house faster?.

More about investment

How to not stutter.

How to be a good manager.

Scrubbing.

How to use Google.

How to talk to strangers.

Potential topics:

How to give clear instructions.

How to see things from other people's point of view

Cool sidetrack: Fish and lightning

Links and quotes:

Why grad school in the humanities is a bad choice One probably could not devise a better system for keeping people with humanistic values away from power than by confining them to decade-long graduate programs with a long future of transient adjunct positions making less than the minimum wage. From Part 2. The first article is a nice example of applying the far view (look at how things are in general) to a personal decision.

Quantified Health Prize results announced

Follow-up to: Announcing the Quantified Health Prize

I am happy to announce that Scott Alexander, better known on Less Wrong as Yvain, has won the first Quantified Health Prize, and Kevin Fischer has been awarded second place. There were exactly five entries, so the remaining prizes will go to Steven Kaas, Kevin Keith and Michael Buck Shlegeris.

The full announcement can be found here until the second contest is announced, and is reproduced below the fold.

While we had hoped to receive more than five entries, I feel strongly that we still got our money's worth and more. Scott Alexander and Kevin Fischer in particular put in a lot of work, and provided largely distinct sets of insight into the question. In general, it is clear that much time was usefully spent, and all five entries had something unique to contribute to the problem.

We consider the first contest a success, and intend to announce a second contest in the next few weeks that will feature multiple questions and a much larger prize pool.

Discussion of all five entries follows:

5th Place ($500): Full report

Steven Kaas makes a well-reasoned argument for selenium supplementation. That obviously wasn't a complete entry. It's very possible this was a strategic decision in the hopes there would be less than five other entries, and if so it was a smart gamble that paid off. I sympathize with his statements on the difficulty of making good decisions in this space.

4th Place ($500): Full report

Kevin Keeth's Recommendation List is as follows: "No quantified recommendations were developed. See 'Report Body' for excruciating confession of abject failure." A failure that is admitted to and analyzed with such honesty is valuable, and I'm glad that Kevin submitted an entry rather than giving up, even though he considered his entry invalid and failure is still failure. Many of the concerns raised in his explanation are no doubt valid concerns. I do think it is worth noting that a Bayesian approach is not at a loss when the data is threadbare, and the probabilistic consequences of actions are highly uncertain. Indeed, this is where a Bayesian approach is most vital, as other methods are forced to throw up their hands. Despite the protests, Kevin does provide strong cases against supplementation of a number of trace minerals that were left unconsidered by other entries, which is good confirmation to have.

3rd Place ($500): Full report

Michael Buck Shlegeris chose to consider only five minerals, but made reasonable choices of which five to include. None of the recommendations made on those five seem unreasonable, but the reasoning leading to them is unsound. This starts with the decision to exclude studies with less than a thousand participants. While larger sample sizes are obviously better (all else being equal), larger studies also tend to be retrospective, longitudinal monitoring studies and meta-analyses. The conclusions in each section are not justified by the sources cited, and the RDI (while a fine starting point) is leaned on too heavily. There is no cost/benefit analysis, nor are the recommendations quantified. This is a serious entry, but one that falls short.

2nd Place ($1000): Full report

Kevin Fischer provides a treasure trove of information, teasing out many fine details that the other entries missed, and presented advocacy of an alternate approach that treats supplementation as a last resort far inferior to dietary modifications. Many concerns were raised about method of intake, ratios of minerals, absorption, and the various forms of each mineral. This is impressive work. There is much here that we will need to address seriously in the future, and we're proud to have added Kevin Fischer to our research team; he has already been of great help, and we will no doubt revisit these questions.

Unfortunately, this entry falls short in several important ways. An important quote from the paper:

""Eat food high in nutrients" represents something like the null hypothesis on nutrition - human beings were eating food for millions of years before extracting individual constituents was even possible. "Take supplements" is the alternative hypothesis.

This is an explicitly frequentist, and also Romantic, approach to the issue. Supplementation can go wrong, but so can whole foods, and there's no reason to presume that what we did, or are currently doing with them, is ideal. Supplementation is considered only as a last resort, after radical dietary interventions have "failed," and no numbers or targets for it are given. No cost-benefit analysis is done on either supplementation or on the main recommendations.

Winner: Scott Alexander / Yvain ($5000): Full report

Scott Alexander's entry was not perfect, but it did a lot of things right. An explicit cost/benefit analysis was given, which was very important. The explanations of the origins of the RDAs were excellent, and overall the analysis of various minerals was strong, although some factors found by Kevin were missed. Two of the analyses raised concerns worth noting: potassium and sodium.

On sodium, the concern was that the analysis treated the case as clear cut when it was not; there have been challenges to salt being bad, as referenced last year by Robin Hanson, and the anti-salt studies are making the two-stage argument that blood pressure causes risks and salt raises blood pressure, rather than looking at mortality. However, the conclusions here are still reasonable, especially for ordinary Americans regularly eating super-stimulus foods loaded with the stuff.