Best of LessWrong: March 2014

1. Strategic choice of identity
2. Optimal Exercise
3. Political Skills which Increase Income
4. A vote against spaced repetition
5. Self-Congratulatory Rationalism
6. Less Wrong Study Hall - Year 1 Retrospective
7. Don't teach people how to reach the top of a hill
8. On not getting a job as an option

Best of LessWrong: March 2014

1. Strategic choice of identity
2. Optimal Exercise
3. Political Skills which Increase Income
4. A vote against spaced repetition
5. Self-Congratulatory Rationalism
6. Less Wrong Study Hall - Year 1 Retrospective
7. Don't teach people how to reach the top of a hill
8. On not getting a job as an option

Strategic choice of identity

Identity is mostly discussed on LW in a cautionary manner: keep your identity small, be aware of the identities you are attached to. As benlandautaylor points out, identities are very powerful, and while being rightfully cautious about them, we can also cultivate them deliberately to help us achieve our goals.

Some helpful identities that I have that seem generally applicable:

- growth mindset
- low-hanging fruit picker
- truth-seeker
- jack-of-all trades (someone who is good at a variety of skills)
- someone who tries new things
- universal curiosity
- mirror (someone who learns other people's skills)

Out of the above, the most useful is probably growth mindset, since it's effectively a meta-identity that allows the other parts of my identity to be fluid. The low-hanging fruit identity helps me be on the lookout for easy optimizations. The universal curiosity identity motivates me to try to understand various systems and fields of knowledge, besides the domains I'm already familiar with. It helps to give these playful or creative names, for example, "champion of low-hanging fruit". Some of these work well together, for example the "trying new things" identity contributes to the "jack of all trades" identity.

It's also important to identify unhelpful identities that get in your way. Negative identities can be vague like "lazy person" or specific like "someone who can't finish a project". With identities, just like with habits, the easiest way to reduce or eliminate a bad one seems to be to install a new one that is incompatible with it. For example, if you have a "shy person" identity, then going to parties or starting conversations with strangers can generate counterexamples for that identity, and help to displace it with a new one of "sociable person". Costly signaling can be used to achieve this - for example, joining a public speaking club. The old identity will not necessarily go away entirely, but the competing identity will create cognitive dissonance, which it can be useful to deliberately focus on. More specific identities require more specific counterexamples. Since the original negative identity makes it difficult to perform the actions that generate counterexamples, there needs to be some form of success spiral that starts with small steps.

Some examples of unhelpful identities I've had in the past were "person who doesn't waste things" and "person with poor intuition". The aversion to wasting money and material things predictably led to wasting time and attention instead. I found it useful to try "thinking like a trader" to counteract this "stingy person" identity, and get comfortable with the idea of trading money for time. Now I no longer obsess about recycling or buy the cheapest version of everything. Underconfidence in my intuition was likely responsible for my tendency to miss the forest for the trees when studying math or statistics, where I focused on details and missed the big picture ideas that are essential to actual understanding. My main objection to intuitions was that they feel imprecise, and I am trying to develop an identity of an "intuition wizard" who can manipulate concepts from a distance without zooming in. That is a cooler name than "someone who thinks about things without really understanding them", and brings to

mind some people I know who have amazing intuition for math, which should help the identity stick.

There can also be ambiguously useful identities, for example I have a "tough person" identity, which motivates me to challenge myself and expand my comfort zone, but also increases self-criticism and self-neglect. Given the mixed effects, I'm not yet sure what to do about this one - maybe I can come up with an identity that only has the positive effects.

Which identities hold you back, and which ones propel you forward? If you managed to diminish negative identities, how did you do it and how far did you get?

Optimal Exercise

Followup to: Lifestyle interventions to increase longevity.

What does it mean for exercise to be optimal?

- Optimal for looks
- Optimal for time
- Optimal for effort
- Optimal for performance
- Optimal for longevity

There may be even more criteria.

We're all likely going for a mix of outcomes, and optimal exercise is going to change depending on your weighting of different factors. So I'm going to discuss something close to a minimum viable routine based on meta-analyses of exercise studies.

Not knowing which sort of exercise yields the best results gives our brains an excuse to stop thinking about it. The intent of this post is to go over the dose responses to various types of exercise. We're going to break through vague notions like "exercise is good" and "I should probably exercise more" with a concrete plan where you understand the relevant parameters that will cause dramatic improvements.

How much exercise?

Optimality aside, I recommend starting with a very minimal routine for 6-ish weeks to build the habit of exercise in to your life. You'll want a program that causes you little mental stress that you can actually stick with. You've got a few options for achieving this. The gains from weightlifting can be surprisingly quick—you'll see dramatic changes in your appearance in 4 months—and seeing yourself lift more weight every session can be a great motivator. Couch to 5k is a basic running progression designed for sedentary people. A daily bodyweight routine is a good way to achieve habit formation through consistency. I recommend making a firm choice and sticking with it until it becomes easy.

Once you've made exercise a habit, you'll want to gradually nudge yourself towards the level that's optimal. So what is that level? Most of the rest of the claims in this post are supported by this review by Swiss researchers. As far as I know, this is the largest systematic review of exercise studies ever undertaken, reviewing 7000 studies with 80 meeting inclusion criteria covering over 1.3 million subjects. Sheer size, however, is not the only reason to take this study very seriously. As someone who has read hundreds of exercise studies, I can say that the methodology of the meta-analysis done to determine dose-response to exercise is excellent. What is most encouraging is that the study authors repeatedly point out shortcomings, and ways their findings should *not* be interpreted because the underlying data does not warrant it. They also check for publication bias. One potential caveat is that this a review of cohort studies, not RCTs. But the authors note that RCTs of exercise almost always show *greater* effect sizes, not smaller. This is likely because people over-report how much exercise they do in observational studies.

In order to compare the intensity of different activities, exercise researchers use a unit called a MET, or metabolic equivalent. The MET is defined so that your weight (in kg) * METs = Calories you're burning per hour. An example MET table can be found here. For the purposes of exercise studies, activities are typically classified as low-intensity, moderate-intensity, and vigorous-intensity. These roughly correspond to 1-3, 4-6, and 7+ METs per hour. For typical individuals, this will translate to approximately 200, 400, and 600+ Calories burned per hour.

On the low end, some studies have found dramatic benefits from just the first 15 minutes of moderate-intensity exercise per week. These studies indicate that you gain about as much going from no exercise to some exercise as you do going from some exercise to optimal exercise.

The Swiss review finds that the first hour per week of vigorous-intensity activity gets you 2/3rds the benefit of 10 hours per week, but the study authors make sure to point out that this is an implausible effect size and that there are almost certainly some confounding and reverse causality issues going on. Which is to say that people who have better health are simply going to be capable of more exercise.

How about on the high end? Studies differ on where the point of diminishing returns is. Some put it at 1000-1500 Calories; others as high as 3500 Calories. (Remember, a typical individual burns ~400 Calories per hour of moderate-intensity exercise.) I'll shoot for 1500 Calories in my recommendations; 3500 Calories is pretty hard to reach without exercising like a pro athlete.

Estimates indicate that each minute of exercise gets you 3-7 minutes of extra life on average, with higher returns for more intense exercise. So every week, you have the opportunity to get a 3-7x ROI on time spent exercising up to the point of diminishing returns. I recommend high-intensity exercise—not only does it save time, it's also been shown to improve health more on a per-Calorie basis.

Which exercises?

Weight training programs

You may have shied away from weight training in the past because you thought you would turn into some huge gross bodybuilder. But bodybuilders and fitness models take drugs and spend years training intensively to look the way they do. You are not going to gain 20lbs of muscle overnight magically. This goes double if you're a woman. You do not have testosterone; you are not going to be building huge muscles no matter what you do.

Of the forms of exercise I cover, weight training has the most rigorous evidence separating what works and what doesn't. This study (pdf warning) examines what sort of resistance training results in the most rapid improvements.

In weight training lingo, AxB means A sets of B repetitions. So 4x10 would mean 40 reps with rest periods every 10 reps. Our study recommends starting with 4x10 3 times a week, and transitioning to 4x4 2 times a week as you become stronger. Aim for a weight you can barely complete all the reps with.

For an efficient full-body workout, select one exercise from each movement pattern:

Upper push: bench press, incline press, overhead press, dips.

Upper pull: cable rows, barbell rows, dumbbell rows, chin-ups, face pulls.

Lower push: squats, lunges, leg press.

Lower pull: deadlifts, power cleans, hyperextensions, romanian deadlifts, reverse hyperextensions, glute-ham raises.

So a good starting routine would be

A: 4x10 each of squat, bench press, lat pulldown, hyperextension

B: 4x10 each of squat, overhead press, cable row, hyperextension

alternating A and B workouts on different days of the week e.g. AxBxAxx, BxAxBxx.

You'll try to increase the weight by 5lbs each session. As you improve, you want to decrease the reps and increase the intensity so you can keep advancing. For example, if you stall a couple times doing 4x10 at 125lbs on your squat, switch to 4x8 and keep increasing the weight, then 4x6, etc. until you get to something close to an optimal trained routine:

A: 4x4 each of squats, bench, weighted chins, deadlifts

B: 4x4 each of squats, overhead press, barbell row, power cleans

At this point, you're going to the gym only twice per week to give yourself more recovery time.

For learning exercises, there are many tutorials available online and I recommend checking some out if you are confused about form. You can always search for "<name of exercise> tutorial" and get articles and Youtube videos. Many people feel silly practicing their form with extremely light weights (often just the empty bar). But many world record holders start EVERY session this way to warm up and cement muscle memory. Others are silly NOT to do this. Also keep an eye on your ego. It's easy when setting goals for yourself to try to lift a weight that you can't really lift with proper form, because you want to set that personal best. But you'll feel pretty stupid when you are forced to miss the gym for a month because you hurt yourself.

On exercise selection: I'm not a big fan of deadlifts for absolute newbies, unlike say Mark Rippetoe in Starting Strength. Maybe add deadlifts in after you've gained some muscle and you have better awareness of form. I also differ from Rippetoe in recommending that newbies high bar squat (the distinction being that low bar squats place the bar across the shoulders and high bar squats place the bar on the trapezius). I have taught newbies both forms and most find high bar squatting easier to figure out how to do properly. I spent months learning to low bar squat and still injured myself; high bar squatting can be taught in a couple sessions in my experience. Pay attention to whether a tutorial video is trying to teach you low bar squatting; the cues for each exercise are different.

Free weights are generally better than machine exercises, but I recommend cable rows and lat pulldowns to newbies. The goal is to move from cable rows to barbell rows, and from lat pulldowns to actual chin-ups. The issue here is that a beginner won't be able to do the requisite sets and reps of chin-ups and rows with good form.

A note about equipment: Weightlifting shoes have an incredibly high return on investment. They make back injuries less likely, and drastically improve subjective experience of squatting. You can get Rogue weightlifting shoes (use your size in men's dress shoes to size them regardless of gender) for around $120; there are cheaper options available but good shoes will last years so the amortized cost is low. Here's a full list of options; note that even the cheapest weightlifting shoes are miles better than lifting in tennis shoes. I don't have any personal experience with the Reebok CrossFit lifter, but they seem like a good option under $100 for a shoe with the desired .75-inch rigid heel. I recommend a cheap belt (expensive ones aren't any better) in order to improve your execution of the valsalva maneuver during squats and deadlifts which further protects the spine from flexion under load.

Bodyweight routines

For a beginner, something like this is reasonable. Of course such a program will max out in fitness gains fairly quickly, even if you start doing several cycles of it. But this isn't our worry as a beginner. For someone serious about progressing with body weight exercises past this stage, I recommend a program like Overcoming Gravity or Building the Gymnastic Body. There is not really formal support for the efficacy of these programs, but they are endorsed by coaches who train many people successfully, and are consistent with the general principles of weight lifting (progressive overload, training frequency, etc.).

Cardio routines

For cardio, I recommend against high-intensity intervals when starting out. High-intensity intervals carry a greater risk of injury, especially if you're not used to them. They're also unpleasant and not conducive to building habits. For starting out, I recommend something that is based more on psychological results rather than performance optimality, like Couch to 5k. As you progress, start adding in short bursts of more intense effort. The idea is to tire yourself out quickly. If your cardio routine lasts more than 30 minutes you're probably going too easy.

What type of cardio should you do? Cardio that is amenable to high intensity is probably one of: running (especially up hills), swimming, rowing, biking, burpees, or jump rope. But you might be able to adapt others. I'm a huge fan of rowing for a few reasons. One, it works more than just the legs. Two, you can have a rowing machine in your house, which drastically lowers activation cost. Three, I just find it less aversive subjectively. You can keep stationary bicycle mounts in your house as well, and they have the advantages of being compact and very cheap if you already own a bike. Burpees require no equipment, but they bothered my knees. They work great for some people though. Jumping rope is also very space/time efficient but the skill required acts as something of a barrier. If you find learning the skill enjoyable, it's a great option. It is worth noting that runners, bikers, and rowers have among the best VO2 max scores of any athletes.

What does the optimal high-intensity cardio routine look like? Data on this comes from this Meta-analysis of VO2 max trainability. VO2 max has been shown to be a robust predictor of mortality. This relation has held across elite athletes, to average individuals, to the overweight (see Figure 2 from this meta-analysis of vo2 max trainability). Unfortunately, it appears that the protocols eliciting the greatest

increases in VO2 max are so arduous as to have high attrition rates. 6 days/week is not a schedule of training I expect anyone but professional athletes to maintain. What sort of realistic routine can still achieve most of these gains? The meta-analysis supports a mixture of 3-5 minute intervals, and longer duration but still intense intervals (30-40 minutes of continuous training). The authors also note that they did not include analysis of the evidence that very high intensity exercise (1 minute or less of max effort) shows unique health benefits. We could simply conclude that a mixture of interval times is good, and that every increment of cardio up to very high levels is likely good for us, but that doesn't feel very motivating. What we want is a clear goal. I'm going to combine data from the VO2 max study and the Swiss review to get a rough estimate. If we want to do resistance training twice per week and cardio at least twice per week can we realistically burn the 1500 calories we want? Let's see. Interval training sessions can vary widely in number of calories burned, but a sprinting session, a 4x4 protocol (one of the more popular protocols in the VO2 max meta analysis, consisting of four 4-minute intervals), and a 30 minute run can burn between 200-450 calories as a first approximation. Twice weekly and we have 400-900 calories. This leaves 600-1100 calories for 2 weightlifting sessions. Estimates for calories burned weightlifting vary extremely widely, most likely due to the huge number of exercises considered "weightlifting", but even the lower estimates put it over 300 calories per hour. This puts 2-4 hours of weightlifting per week at 600-1200 calories expended. How convenient!

All that remains is to suggest specific cardio routines. I don't have the evidence to say with any confidence what a truly optimal routine would look like here, but I can at least give well studied examples of each.

Very high intensity routines follow a pattern of a short warmup (5 minutes at a slow pace) followed by several bursts of 10-60 seconds all out intensity. (30 on 30 off for 10 intervals is popular and close to maximizing vV02max)

VO2 max interval training consists of four 3-5 minute intervals at 85%-95% your max heart rate interspersed with slower jogging for the same interval.

Longer interval training consists of 20-40 minute runs at a consistent pace such that you are exhausted by the end.

I wouldn't worry about the optimal frequency for each one. Don't forget that even training populations just consistently doing one type shows very dramatic improvements in health. I'd suggest freely mixing them up and trying to have fun with it.

Summary of my recommended routine

This is what I recommend gradually working towards once you've made exercise a habit:

- ~1-2 hour weightlifting sessions 2-3x a week. (A third weightlifting session is recommended for the first several months, for both gaining strength and building habits.)
- ~15-40 minutes of vigorous cardio 2-3x a week.

Don't do vigorous cardio on the same day as lifting weights! It's a good way to injure yourself, especially your lower back. Exercise doesn't make you stronger; it makes you

weaker. It's the recovery from exercise that makes you stronger; give your body time to recover.

Nutrition

Don't try to implement a new diet and a new exercise plan at the same time. If you're trying to choose, do an exercise plan first—effects on health are much larger.

If you are underweight or normal weight, you'll need to eat more when you start exercising. Celebrate after your workouts by eating to reinforce the exercise habit. You may think eating pizza is bad for you, but not exercising is worse, so reward yourself however you want. Or drink my nutrient dense shake, designed to be consumed after workouts. (John_Maxwell_IV and I are planning to commercialize it after we roll out our first nutritionally complete food.)

If you're overweight: I agree with Gary Taubes that exercise is NOT a good way to lose weight. But exercise has bigger effects on health than weight loss, so I actually recommend prioritizing exercise over changing your diet. (Like I said, don't try to do both at once.)

Note that you don't need to stuff yourself with massive amounts of protein to build muscle. Studies have never shown a measured benefit to consumption above .64g/lb of bodyweight, which translates to around 100g for a 150-160lb person. A single serving (3oz) of chicken, for example, contains about 21g of protein.

If you've made it this far, congratulations; you are now as knowledgeable as any personal trainer I've spoken with.

Political Skills which Increase Income

Summary: This article is intended for those who are "earning to give" (i.e. maximize income so that it can be donated to charity). It is basically an annotated bibliography of a few recent meta-analyses of predictors of income.

Key Results

- *The degree to which management "sponsors" your career development is an important predictor of your salary, as is how skilled you are politically.*
- *Despite the stereotype of a silver-tongued salesman preying on people's biases, rational appeals are generally the best tactic.*
- *After rationality, the best tactics are types of ingratiation, including flattery and acting modest.*

Ng et al. performed a metastudy of over 200 individual studies of objective and subjective career success. Here are the variables they found best correlated with salary:

Predictor	**Correlation**
Political Knowledge & Skills	0.29
Education Level	0.29
Cognitive Ability (as measured by standardized tests)	0.27
Age	0.26
Training and Skill Development Opportunities	0.24
Hours Worked	0.24
Career Sponsorship	0.22

(all significant at p = .05)

(For reference, the "Big 5" personality traits all have a correlation under 0.12.)

Before we go on, a few caveats: while these correlations are significant and important, none are overwhelming (the authors cite Cohen as saying the range 0.24-0.36 is "medium" and correlations over 0.37 are "large"). Also, in addition to the usual correlation/causation concerns, there is lots of cross-correlation: e.g. older people might have greater political knowledge but less education, thereby confusing things. For a discussion of moderating variables, see the paper itself.

Career Sponsorship

There are two broad models of career advancement: contest-mobility and sponsorship-mobility. They are best illustrated with an example.

Suppose Peter and Penelope are both equally talented entry-level employees. Under the contest-mobility model, they would both be equally likely to get a raise or promotion, because they are equally skilled.

Sponsorship-mobility theorists argue that even if Peter and Penelope are equally talented, it's likely that one of them will catch the eye of senior management. Perhaps it's due to one of them having an early success by chance, making a joke in a meeting, or simply just having a more memorable name, like Penelope. This person will be singled out for additional training and job opportunities. Because of this, they'll have greater success in the company, which will lead to more opportunities etc. As a result, their initial small discrepancy in attention gets multiplied into a large differential.

The authors of the metastudy found that self-reported sponsorship levels (i.e. how much you feel the management of your company "sponsors" you) have a significant, although moderate, relationship to salary. Therefore, the level at which you currently feel sponsored in your job should be a factor when you consider alternate opportunities.

The Dilbert Effect

The strongest predictor of salary (tied with education level) is what the authors politely term “Political Knowledge & Skills” - less politely, how good you are at manipulating others.

Several popular books (such as Cialdini’s *Influence*) on the subject of influencing others exist, and the study of these “influence tactics” in business stretches back 30 years to Kipnis, Schmidt and Wilkinson. Recently, Higgins et al. reviewed 23 individual studies of these tactics and how they relate to career success. Their results:

Tactic	**Correlation**	**Definition (From Higgins et al.)**
Rationality	0.26	Using data and information to make a logical argument supporting one's request
Ingratiation	0.23	Using behaviors designed to increase the target's liking of oneself or to make oneself appear friendly in order to get what one wants
Upward Appeal	0.05	Relying on the chain of command, calling in superiors to help get one's way
Self-Promotion	0.01	Attempting to create an appearance of competence or that you are capable of completing a task
Assertiveness	-0.02	Using a forceful manner to get what one wants
Exchange	-0.03	Making an explicit offer to do something for another in exchange for their doing what one wants

(Only ingratiation and rationality are significant.)

This site has a lot of information on how to make rational appeals, so I will focus on the less-talked-about ingratiation techniques.

How to be Ingratiating

Gordon analyzed 69 studies of ingratiation and found the following. (Unlike the previous two sections, success here is measured in lab tests as well as in career advancement. However, similar but less comprehensive results have been found in terms of career success):

Tactic	**Weighted Effectiveness (Cohen’s d difference between control and intervention)**	**Description**
Other Enhancement	0.31	Flattery
Opinion Conformity	0.23	“Go along to get along”
Self-presentation	0.15	Any of the following tactics: Self-promotion, self-deprecation, apologies, positive nonverbal displays and name usage
Combination	0.10	Includes studies where the participants weren’t told which strategy to use, in addition to when they were instructed to use multiple strategies
Rendering Favors	0.05	

Self-presentation is split further:

Tactic	Weighted Effect Size	Comment
Modesty	0.77	
Apology	0.59	Apologizing for poor performance
Generic	0.28	When the participant is told in generic terms to improve their self-presentation
Nonverbal behavior and name usage	-0.14	Nonverbal behavior includes things like wearing perfume. Name usage means referring to people by name instead of a pronoun.
Self-promotion	-0.17	

Moderators

One important moderator is the direction of the appeal. If you are talking to your boss, your tactics should be different than if you're talking to a subordinate. Other-enhancement (flattery) is always the best tactic no matter who you're talking to, but when talking to superiors it's by far the best. When talking to those at similar levels to you, opinion conformity comes close to flattery, and the other techniques aren't far behind.

Unsurprisingly, when the target realizes you're being ingratiating, the tactic is less effective. (Although effectiveness doesn't go to zero - even when people realize you're flattering them just to suck up, they generally still appreciate it.) Also, women are better at being ingratiating than men, and men are more influenced by these ingratiating tactics than women. The most important caveat is that lab studies find much larger effect sizes than in the field, to the extent that the average field effect for the ingratiating tactics is negative. This is probably due to the fact that lab experiments can be better controlled.

Conclusion

It's unlikely that a silver-tongued receptionist will out-earn an introverted engineer. But simple techniques like flattery and attempting to get "sponsored" can appreciably improve returns, to the extent that political skills are one of the strongest predictors of salaries.

I would like to thank Brian Tomasik and Gina Stuessy for reading early drafts of this article.

References

Cohen, Jacob. *Statistical power analysis for the behavioral sciences*. Psychology Press, 1988.

Gordon, Randall A. "Impact of ingratiation on judgments and evaluations: A meta-analytic investigation." *Journal of Personality and Social Psychology* 71.1 (1996): 54.

Higgins, Chad A., Timothy A. Judge, and Gerald R. Ferris. "Influence tactics and work outcomes: a meta-analysis." *Journal of Organizational Behavior* 24.1 (2003): 89-106.

Judge, Timothy A., and Robert D. Bretz Jr. "Political influence behavior and career success." *Journal of Management* 20.1 (1994): 43-65.

Kipnis, David, Stuart M. Schmidt, and Ian Wilkinson. "Intraorganizational influence tactics: Explorations in getting one's way." *Journal of Applied psychology* 65.4 (1980): 440.

Ng, Thomas WH, et al. "Predictors of objective and subjective career success: A meta-analysis." *Personnel psychology* 58.2 (2005): 367-408.

A vote against spaced repetition

LessWrong seems to be a big fan of spaced-repetition flashcard programs like Anki, Supermemo, or Mnemosyne. I used to be. After using them religiously for 3 years in medical school, I now categorically advise against using them for large volumes of memorization.

[A caveat before people get upset: I think they appropriate in certain situations, and I have not tried to use them to learn a language, which seems its most popular use. More at the bottom.]

A bit more history: I and 30 other students tried using Mnemosyne (and some used Anki) for multiple tests. At my school, we have a test approximately every 3 weeks, and each test covers about 75 pages of high-density outline-format notes. Many stopped after 5 or so such tests, citing that they simply did not get enough returns from their time. I stuck with it longer and used them more than anyone else, using them for 3 years.

Incidentally, I failed my first year and had to repeat.

By the end of that third year (and studying for my Step 1 boards, a several-month process), I lost faith in spaced-repetition cards as an effective tool for my memorization demands. I later met with a learning-skills specialist, who felt the same way, and had better reasons than my intuition/trial-and-error:

- Flashcards are less useful to learning the “big picture”
- Specifically, if you are memorizing a large amount of information, there is often a hierarchy, organization, etc that can make leaning the whole thing easier, and you loose the constant visual reminder of the larger context when using flashcards.
- Flashcards do not take advantage of spatial, mapping, or visual memory, all of which the human mind is much better optimized for. It is not so well built to memorize pairs between seemingly arbitrary concepts with few to no intuitive links. My preferred methods are, in essence, hacks that use your visual and spatial memory rather than rote.

Here are examples of the typical kind of things I memorize every day and have found flashcards to be surprisingly worthless for:

- The definition of Sjögren's syndrome
- The contraindications of Metronidazole
- The significance of a rise in serum αFP

Here is what I now use in place of flashcards :

1. Ven diagrams/etc, to compare and contrast similar lists. (This is more specific to medical school, when you learn subtly different diseases.)
2. Mnemonic pictures. I have used this myself for years to great effect, and later learned it was taught by my study-skills expert, though I'm surprised I haven't found them formally named and taught anywhere else. The basic concept is to make a large picture, where each detail on the picture corresponds to a detail you want to memorize.

3. Memory palaces. I recently learned how to properly use these, and I'm a true believer. When I only had the general idea to “pair things you want to memorize with places in your room” I found it worthless, but after I was taught a lot of do's and don'ts, they're now my favorite way to memorize any list of 5+ items. If there's enough demand on LW I can write up a summary.

Spaced repetition is still good for knowledge you need to retrieve *immediately*, when a 2-second delay would make it useless. I would still consider spaced-repetition to memorize some of the more rarely-used notes on the treble and bass clef, if I ever decide to learn to sight-read music properly. I make no comment on it's usefulness to learn a foreign language, as I haven't tried it, but if I were to pick one up I personally would start with a rosetta-stone-esque program.

Your mileage may vary, but after seeing so many people try and reject them, I figured it was enough data to share. Mnemonic pictures and memory palaces are slightly time consuming when you're learning them. However, if someone has the motivation and discipline to make a stack of flashcards and study them every day indefinitely, then I believe learning and using those skills is a far better use of time.

Self-Congratulatory Rationalism

Quite a few people complain about the atheist/skeptic/rationalist communities being self-congratulatory. I used to dismiss this as a sign of people's unwillingness to admit that rejecting religion, or astrology, or whatever, was any more rational than accepting those things. Lately, though, I've started to worry.

Frankly, there seem to be a *lot* of people in the LessWrong community who imagine themselves to be, not just more rational than average, but paragons of rationality who other people should accept as such. I've encountered people talking as if it's ridiculous to suggest they might sometimes respond badly to being told the truth about certain subjects. I've encountered people asserting the rational superiority of themselves and others in the community for flimsy reasons, or no reason at all.

Yet the readiness of members of the LessWrong community to disagree with and criticize each other suggests we don't actually think all that highly of each other's rationality. The fact that members of the LessWrong community tend to be smart is no guarantee that they will be rational. And we have much reason to fear "rationality" degenerating into signaling games.

What Disagreement Signifies

Let's start by talking about disagreement. There's been a lot of discussion of disagreement on LessWrong, and in particular of Aumann's agreement theorem, often glossed as something like "two rationalists can't agree to disagree." (Or perhaps that we can't foresee to disagree.) Discussion of disagreement, however, tends to focus on what to do about it. I'd rather take a step back, and look at what disagreement tells us about ourselves: namely, that we don't think all that highly of each other's rationality.

This, for me, is the take-away from Tyler Cowen and Robin Hanson's paper Are Disagreements Honest? In the paper, Cowen and Hanson define honest disagreement as meaning that "meaning that the disputants respect each other's relevant abilities, and consider each person's stated opinion to be his best estimate of the truth, given his information and effort," and they argue disagreements aren't honest in this sense.

I don't find this conclusion surprising. In fact, I suspect that while people sometimes *do* mean it when they talk about respectful disagreement, often they realize this is a polite fiction (which isn't necessarily a bad thing). Deep down, they know that disagreement is disrespect, at least in the sense of not thinking that highly of the other person's rationality. That people know this is shown in the fact that they don't like being told they're wrong—the reason why Dale Carnegie says you can't win an argument.

On LessWrong, people are quick to criticize each others' views, so much so that I've heard people cite this as a reason to be reluctant to post/comment (again showing they know intuitively that disagreement is disrespect). Furthermore when people in LessWrong criticize others' views, they very often don't seem to expect to quickly reach agreement. Even people Yvain would classify as "experienced rationalists" sometimes knowingly have persistent disagreements. This suggests that LessWrongers almost never consider each other to be perfect rationalists.

And I actually think this is a sensible stance. For one thing, even if you met a perfect rationalist, it could be hard to figure out that they are one. Furthermore, the problem of knowing what to do about disagreement is made harder when you're faced with other people having persistent disagreements: if you find yourself agreeing with Alice, you'll have to think Bob is being irrational, and vice versa. If you rate them equally rational and adopt an intermediate view, you'll have to think they're both being a bit irrational for not doing likewise.

The situation is similar to Moore's paradox in philosophy—the impossibility of asserting "it's raining, but I don't believe it's raining." Or, as you might say, "Of course I think my opinions are right and other people's are wrong. Otherwise I'd change my mind." Similarly, when we think about disagreement, it seems like we're forced to say, "Of course I think my opinions are rational and other people's are irrational. Otherwise I'd change my mind."

We can find some room for humility in an analog of the preface paradox, the fact that the author of a book can say things like "any errors that remain are mine." We can say this because we might think each individual claim in the book is highly probable, while recognize that all the little uncertainties add up to it being likely there are still errors. Similarly, we can think each of our beliefs are individually rational, while recognizing we still probably have *some* irrational beliefs—we just don't know which ones And just because respectful disagreement is a polite fiction doesn't mean we should abandon it.

I don't have a clear sense of how controversial the above will be. Maybe we all already recognize that we don't respect each other's opinions 'round these parts. But I think some features of discussion at LessWrong look odd in light of the above points about disagreement—including some of the things people say about disagreement.

The wiki, for example, says that "Outside of well-functioning prediction markets, Aumann agreement can probably only be approximated by careful deliberative discourse. Thus, fostering effective deliberation should be seen as a key goal of Less Wrong." The point of Aumann's agreement theorem, though, is precisely that ideal rationalists *shouldn't* need to engage in deliberative discourse, as usually conceived, in order to reach agreement.

As Cowen and Hanson put it, "Merely knowing someone else's opinion provides a powerful summary of everything that person knows, powerful enough to eliminate any differences of opinion due to differing information." So sharing evidence the normal way shouldn't be necessary. Asking someone "what's the evidence for that?" implicitly says, "I don't trust your rationality enough to take your word for it." But when dealing with real people who may or may not have a rational basis for their beliefs, that's almost always the right stance to take.

Intelligence and Rationality

Intelligence does not equal rationality. Need I say more? Not long ago, I wouldn't have thought so. I would have thought it was a fundamental premise behind LessWrong, indeed behind old-school scientific skepticism. As Michael Shermer once said, "Smart people believe weird things because they are skilled at defending beliefs they arrived at for non-smart reasons."

Yet I've heard people suggest that you must never be dismissive of things said by smart people, or that the purportedly high IQ of the LessWrong community means people here don't make bad arguments. When I hear that, I think "*whaaat?* People on LessWrong make bad arguments all the time!" When this happens, I *generally* limit myself to trying to point out the flaw in the argument and/or downvoting, and resist the urge to shout "YOUR ARGUMENTS ARE BAD AND YOU SHOULD FEEL BAD." I just think it.

When I reach for an explanation of why terrible arguments from smart people shouldn't surprise anyone, I go to Yvain's Intellectual Hipsters and Meta-Contarianism, one of my favorite LessWrong posts of all time. While Yvain notes that meta-contrarianism often isn't a good thing, though, on re-reading it I noticed what seems like an important oversight:

> A person who is somewhat upper-class will conspicuously signal eir wealth by buying difficult-to-obtain goods. A person who is very upper-class will conspicuously signal that ey feels no need to conspicuously signal eir wealth, by deliberately not buying difficult-to-obtain goods.
>
> A person who is somewhat intelligent will conspicuously signal eir intelligence by holding difficult-to-understand opinions. A person who is very intelligent will conspicuously signal that ey feels no need to conspicuously signal eir intelligence, by deliberately not holding difficult-to-understand opinions.
>
> According to the survey, the average IQ on this site is around 145. People on this site differ from the mainstream in that they are more willing to say death is bad, more willing to say that science, capitalism, and the like are good, and less willing to say that there's some deep philosophical sense in which 1+1 = 3. That suggests people around that level of intelligence have reached the point where they no longer feel it necessary to differentiate themselves from the sort of people who aren't smart enough to understand that there might be side benefits to death.

The pattern of countersignaling Yvain describes here is real. But it's important not to forget that sometimes, the super-wealthy signal their wealth by buying things even the moderately wealthy can't afford. And sometimes, the very intelligent signal their intelligence by holding opinions even the moderately intelligent have trouble understanding. You also get hybrid status moves: designer versions of normally low-class clothes, complicated justifications for opinions normally found among the uneducated.

Robin Hanson has argued that this leads to biases in academia:

> I've argued that the main social function of academia is to let students, patrons, readers, etc. affiliate with credentialed-as-impressive minds. If so, academic beliefs are secondary – the important thing is to clearly show respect to those who make impressive displays like theorems or difficult data analysis. And the obvious way for academics to use their beliefs to show respect for impressive folks is to have academic beliefs track the most impressive recent academic work.

Robin's post focuses on economics, but I suspect the problem is even worse in my home field of philosophy. As I've written before, the problem is that in philosophy, philosophers never agree on whether a philosopher has solved a problem. Therefore, there can be no rewards for being right, only rewards for showing off your impressive intellect. This often means finding clever ways to be wrong.

I need to emphasize that I really do think philosophers are showing off real intelligence, not merely showing off faux-cleverness. GRE scores suggest philosophers are among the smartest academics, and their performance is arguably made more impressive by the fact that GRE quant scores are bimodally distributed based on whether your major required you to spend four years practicing your high school math, with philosophy being one of the majors that doesn't grant that advantage. Based on this, if you think it's wrong to dismiss the views of high-IQ people, you shouldn't be dismissive of mainstream philosophy. But in fact I think LessWrong's oft-noticed dismissiveness of mainstream philosophy is largely justified.

I've found philosophy of religion in particular to be a goldmine of terrible arguments made by smart people. Consider Alvin Plantinga's modal ontological argument. The argument is sufficiently difficult to understand that I won't try to explain it here. If you want to understand it, I'm not sure what to tell you except to maybe read Plantinga's book *The Nature of Necessity*. In fact, I predict at least one LessWronger will comment on this thread with an incorrect explanation or criticism of the argument. Which is not to say they wouldn't be smart enough to understand it, just that it might take them a few iterations of getting it wrong to finally get it right. And coming up with an argument like that is no mean feat—I'd guess Plantinga's IQ is just as high as the average LessWronger's.

Once you understand the modal ontological argument, though, it quickly becomes obvious that Plantinga's logic works just as well to "prove" that it's a necessary truth that pigs fly. Or that Plantinga's god does not exist. Or even as a general purpose "proof" of any purported mathematical truth you please. The main point is that Plantinga's argument is not stupid in the sense of being something you'd only come up with if you had a low IQ—the opposite is true. But Plantinga's argument *is* stupid in the sense of being something you'd only come up with it while under the influence of some serious motivated reasoning.

The modal ontological argument is admittedly an extreme case. Rarely is the chasm between the difficulty of the concepts underlying an argument, and the argument's actual merits, so vast. Still, beware the temptation to affiliate with smart people by taking everything they say seriously.

Edited to add: in the original post, I intended but forgot to emphasize that I think the correlation between IQ and rationality is weak *at best.* Do people disagree? Does anyone want to go out on a limb and say, "They aren't the same thing, but the correlation is still very strong?"

The Principle of Charity

I've made no secret of the fact that I'm not a big fan of the principle of charity—often defined as the rule that you should interpret other people's arguments on the assumption that they are not saying anything stupid. The problem with this is that other people are often saying something stupid. Because of that, I think charitable is over-rated compared to fair and accurate reading. When someone says something stupid, you don't have to pretend otherwise, but it's really important not to attribute to people stupid things they never said.

More frustrating than this simple disagreement over charity, though, is when people who invoke the principle of charity do so selectively. They apply it to people who's views they're at least somewhat sympathetic to, but when they find someone they

want to attack, they have trouble meeting basic standards of fairness. And in the most frustrating cases, this gets explicit justification: "we need to read *these* people charitably, because they are obviously very intelligent and rational." I once had a member of the LessWrong community actually tell me, "You need to interpret me more charitably, because you know I'm sane." "Actually, buddy, I don't know that," I wanted to reply—but didn't, because that would've been rude.

I can see benefits to the principle of charity. It helps avoid flame wars, and from a Machiavellian point of view it's nice to close off the "what I actually meant was..." responses. Whatever its merits, though, they can't depend on the actual intelligence and rationality of the person making an argument. Not only is intelligence no guarantee against making bad arguments, the whole reason we demand other people tell us their reasons for their opinions in the first place is we fear their reasons *might* be bad ones.

As I've already explained, there's a difficult problem here about how to be appropriately modest about our own rationality. When I say something, I never think it's stupid, otherwise I wouldn't say it. But at least I'm not so arrogant as to go around demanding other people acknowledge my highly advanced rationality. I don't demand that they accept "Chris isn't saying anything stupid" as an axiom in order to engage with me.

Beware Weirdness for Weirdness' Sake

There's a theory in the psychology and sociology of religion that the purpose of seemingly foolish rituals like circumcision and snake-handling is to provide a costly and therefore hard-to-fake signal of group commitment. I think I've heard it suggested—though I can't find by who—that crazy religious doctrines could serve a similar purpose. It's easy to say you believe in a god, but being willing to risk ridicule by saying you believe in one god who is three persons, who are all the same god, yet not identical to each other, and you can't explain how that is but it's a mystery you accept on faith... now *that* takes dedication.

Once you notice the general "signal group commitment in costly ways" strategy, it seems to crop up everywhere. Subcultures often seem to go out of their way to be weird, to do things that will shock people outside the subculture, ranging from tattoos and weird clothing to coming up with reasons why things regarded as normal and innocuous in the broader culture are actually evil. Even something as simple as a large body of jargon and in-jokes can do the trick: if someone takes the time to learn all the jargon and in-jokes, you know they're committed.

This tendency is probably harmless when done with humor and self-awareness, but it's more worrisome when a group becomes convinced its little bits of weirdness for weirdness' sake are a sign of its superiority to other groups. And it's worth being aware of, because it makes sense of signaling moves that aren't straightforwardly plays for higher status.

The LessWrong community has amassed a truly impressive store of jargon and in-jokes over the years, and some of it's quite useful (I reiterate my love for the term "meta-contrarian"). But as with all jargon, LessWrongian jargon is often just a silly way

of saying things you could have said without it. For example, people say "I have a poor mental model of..." when they could have just said they don't understand it very well.

That bit of LessWrong jargon is merely silly. Worse, I think, is the jargon around politics. Recently, a friend gave "they avoid blue-green politics" as a reason LessWrongians are more rational than other people. It took a day before it clicked that "blue-green politics" here basically just meant "partisanship." But complaining about partisanship is old hat—literally. America's founders were fretting about it back in the 18th century. Nowadays, such worries are something you expect to hear from boringly middle-brow columnists at major newspapers, not edgy contrarians.

But "blue-green politics," "politics is the mind-killer"... never mind how much content they add, the point is they're obscure enough to work as an excuse to feel superior to anyone whose political views are too mainstream. Outsiders will probably think you're weird, invoking obscure jargon to quickly dismiss ideas that seem plausible to them, but on the upside you'll get to bond with members of your in-group over your feelings of superiority.

A More Humble Rationalism?

I feel like I should wrap up with some advice. Unfortunately, this post was motivated by problems I'd seen, not my having thought of brilliant solutions to them. So I'll limit myself to some fairly boring, non-brilliant advice.

First, yes, some claims are more rational than others. Some people even do better at rationality overall than others. But the idea of a real person being anything close to an ideal rationalist is an extraordinary claim, and should be met with appropriate skepticism and demands for evidence. Don't forget that.

Also, beware signaling games. A good dose of Hansonian cynicism, applied to your own in-group, is healthy. Somewhat relatedly, I've begun to wonder if "rationalism" is really good branding for a movement. Rationality is systematized winning, sure, but the "rationality" branding isn't as good for keeping that front and center, especially compared to, say the effective altruism meme. It's just a little too easy to forget where "rationality" is supposed to connect with the real world, increasing the temptation for "rationality" to spiral off into signaling games.

Less Wrong Study Hall - Year 1 Retrospective

Some time back, a small group of Less Wrongers collected in a video chatroom to work on…things. We've been at it for exactly one year as of today, and it seems like a good time to see what's come of it.[1] So here is what we've done, what we're doing, and a few thoughts on where we're going. At the end is a survey taken of the LWSH, partly to be compared to Less Wrong proper, but mostly for fun. If you like what you see here, come join us. The password is "lw".

A Brief History of the Hall

I think the first inspiration was Eliezer looking for someone to sit with him while he worked, to help with productivity and akrasia. Shannon Friedman answered the call and it seemed to be effective. She suggested a similar coworking scheme to one of her clients, Mqrius, to help him with akratic issues surrounding his thesis. She posted on Less Wrong about it, with the intent of connecting him and possibly others who wanted to co-work in a similar fashion. Tsakinis, in the comments, took the idea a step further, and created a Tinychat video chatroom for group working. It was titled the Less Wrong Study Hall. The theory is that it will help us actually do the work, instead of, say, reading tvtropes when we should be studying. It turned out to be a decent Schelling point, enough to form a regular group and occasionally attract new people. It's grown slowly but steadily.

Tinychat's software sucks, and there have been a couple of efforts to replace it. Mqrius looked into OpenMeetings, but it didn't work out. Yours truly took a crack at programming a LWSH Google Hangout, but it ran aground on technical difficulties. Meanwhile the tinychat room continued to work, and despite nobody actually liking it, it's done the job well enough.

Tinychat is publicly available, and there have been occasional issues with the public along the way. A few people took up modding, but it was still a nuisance. Eventually a password was placed on the room, which mostly shut down the problem. We did have one guy straight out guess the password, which was a…peculiar experience. He was notably not all there, but somehow still scrupulously polite, and left when asked. I don't think I've ever seen that happen on the Internet before.

A year after the Hall opened, we have about twenty to twenty-five regulars, with an unknown number of occasional users. We're still well within Dunbar's number, so everybody knows everybody else and new users integrate quickly. We've developed a reasonably firm set of social norms to guide our work, in spite of not having direct technical control nor clear leaders.

The story thus far has been a pretty good one. I'm looking forward to Year Two.

Current MO and Social Norms

No one went out of their way to come up with these and nobody is really enforcing them, but it's how the group currently behaves.

We adopted the Pomodoro Technique early on to organize our work. We run 32-minute pomodoros instead of the usual 25 (I have no idea how this custom originated; I'm told it was Dentin's idea though), interspersed with eight minute breaks that tend to run over as often as not. Sometimes, we'll run a longer or shorter pomo if someone requests it. When you enter the room, ask for the current time. Usually at least one person will notice and respond with the endpoint of the current pomo, expressed in minutes-past-the-hour to be applicable across time zones.[2]

Don't talk during pomos without good cause. Asking for the current time is good cause; so is saying a brief goodbye when you leave. If someone else is talking enough to be distracting, it's acceptable to ask them to save it for the next break. Otherwise, insert nose in grindstone.

During breaks, talk as much as you want. Talking about what you're working on is explicitly encouraged. Bragging about something you've just completed is also explicitly encouraged. In the absence of those two things, though, it's okay to just be social.[3]

Anyone who notices that the break has ended can start the next pomo just by stating its length and end time. Usually one or two self-selected people will keep the rest of the room on track in this regard. If the break runs over and you're the first to notice, speak up even if you don't want to be responsible for the clock; it's good form to cut the chitchat quickly once someone has noted that it's time to get back to work.

Most of us keep our cameras on in the room, and I encourage new users to do the same. Having someone watch you work seems to improve motivation and reduce the temptation to slack. Your presence will be more obvious, which helps if you happen to be new and are scrolled off the bottom of the user list. It also encourages others to interact with you; socialization during breaks is easier and works better when you can see people's expressions. As noted in the survey below, that sort of social reinforcement is a big plus. In lieu of a camera, some users display their working desktop instead, but this is rare and seems to require a third party program. If you like privacy, keeping everything off is okay too; just be aware that you're avoiding a useful tool.

Bullet point version of the above:

- Say hello and ask for the current time when entering.
- Don't talk during pomos.
- Do talk during breaks.
- Talking about work is encouraged.
- Bragging about work is encouraged.
- Don't turn your mic on.
- But do turn your camera or desktop view on if you want.

Personal Reflections

It's been an interesting first year. I'm part of a tribe as it begins to cohere, and I like it.

Is it effective? Good question. I enjoy the environment but that doesn't mean it's succeeding at what it's supposed to do: help akratics Get Stuff Done. I *feel* like it's effective, but being sure of that is hard. I've certainly written more in the last year than the previous ten. I do use a couple of other productivity hacks, especially

Beeminder, and in fact I frequently work on Beeminded tasks in the Hall. Someone (jkadlubo?) mentioned that they didn't like Beeminder because it was all stick and no carrot. For me, the study hall is the carrot. I *like* working alongside other people.

We have a lot of students, which is unsurprising given that we splintered off Less Wrong and Less Wrong also has a lot of students. This makes me wonder what will happen when our users begin to graduate.

Tinychat, as a technology, sucks. Unfortunately it's the best existing software that does what we want, by virtue of being the *only* existing software that does what we want (Google Hangouts, the most obvious alternative, does not work for our purposes as mentioned above), and at least two attempts to replace it have failed. I believe this is because, while Tinychat is annoying, it's not annoying *enough* for anyone to put in the relatively large effort that seems to be required to replace it – especially since the room is self-selected for akratic problems in the first place. The most promising alternative tech I can find is XMPP Muji, a standard for multi-user video chat services, but the standard's approval is deferred and I'm not sure why. A cursory Google search does not turn up any software that implements it.

One of the problems brought up when the room was first proposed was that people might end up using the Hall to socialize rather than work. The concern was justified, at least partially; we routinely run over breaks talking. But we seem to have countered that with a fairly strong norm of stopping the break as soon as someone notices. As we've all gotten to know each other better, the atmosphere during the breaks has become quite social. I would blame our break-overruns on this, except I'm pretty sure that has been happening since the start. Nevertheless there has been some concern that the more-social atmosphere has been distracting people from actually working. I asked about this in the survey, but the answers were heavily slanted towards "not terribly distracting." (no higher than a 3 on a 1–5 scale, and only three 3's) This leads me to believe that there are more people concerned that others might be bothered than there are others who are actually bothered. It is possible there is a selection effect whereby the people irritated leave and never took the survey; but the change in atmosphere has been relatively recent, and I haven't noticed an exodus of regulars. Additionally, the single largest draw reported is social reinforcement for working. I think the social atmosphere is probably a net positive, although it would be good if we could technologically enforce break limits. Several other people noted that in the suggestion box; unfortunately it is not possible without replacing Tinychat. Which was also a common item in the suggestion box.

We've lost some users along the way (including the original creator, I think), but a near-majority of the room has been around since almost the beginning. That seems to suggest that people who try it out for more than a few cursory sessions stick around. We'll begin to see if that's the case next year, with repeat survey-takers.

Survey and Results

Yvain's Less Wrong 2013 Survey results came out as the one year mark was approaching, and it seemed to me that it would be fun, if maybe not all that useful, to collect the same sort of information about the Hall. So, I did. The survey was open for about three weeks and publicized during Hall breaks by myself and a few other people. You can see the questions used here. I flagrantly plagiarized many of the questions from Yvain's survey (with permission, and thank you), because it seemed like comparing the Hall to Less Wrong proper would be interesting. I took out a few

sections that I didn’t feel were relevant and added a couple of Study Hall specific sections. There is also some silliness at the end.

We had 23 responses. I counted 22 unique users in the room over the period the survey was open. Assuming there are at least some people I did not see during that time (owing to differing schedules) I would guess our actual recurring population is around 30, of which perhaps 20 show up regularly. I have noticed an influx of new faces in the last week or two; I sort of wish I’d run this a month later.

With such a small population it would be fairly easy for community members to de-anonymize people. Respondents were warned of that, but I would still appreciate it if people didn’t go out of their way to try and connect responses to names.

The results are below. My comments are in brackets.

Demographics

Age

Mean	23.4
Stddev	4.2
Min	17
1st quartile	20.5
2nd quartile	23
3rd quartile	26.5
Max	32

[[*Our youngest user is 17 and our oldest is 32, at least among respondents. I feel old now.*]]

Country

Australia	2	9%
Canada	1	4%
Finland	0	0%
France	0	0%
Germany	8	35%
Israel	0	0%
New Zealand	0	0%
Poland	2	9%
Russia	0	0%
United Kingdom	1	4%
United States	4	17%
Other	5	22%

[[*I don’t know why we’re so hugely concentrated in Germany. I think it might be one cluster of people all referring each other.*]]

Race

Asian (East Asian)	0	0%

Asian (Indian Subcontinent)	0	0%
Middle Eastern	0	0%
Black	0	0%
White (non-Hispanic)	21	100%
White (Hispanic)	0	0%
Other	0	0%

[[*I'm* sure *I've seen at least one non-white person…but still.*]]

Sex, Gender, Relationships

I considered not including this section as irrelevant and possibly touchy, especially given how easy it would be to de-anonymize people. Someone (I think tkadlubo) pointed out that Less Wrongers in general are interested in relationship hacking. I ended up making it optional, but everybody answered it anyway.

Sex

Male	18	78%
Female	5	22%
Other	0	0%

Gender

Male (cisgender)	18	78%
Female (cisgender)	5	22%
Male (transgender f->m)	0	0%
Female (transgender m->f)	0	0%
Other	0	0%

Sexual Orientation

Heterosexual	16	70%
Homosexual	0	0%
Bisexual	2	9%
Asexual	1	4%
Other	4	17%

Relationship Style

Prefer monogamous	5	22%
Prefer polyamorous	8	35%
Uncertain / no preference	10	43%
Other	0	0%

[[*All but two of our polyamorous folks were from Europe, and one of those two was me. Clearly I need to move to Europe.*]]

Number of Current Partners

0	13	62%

1	3	14%
2	5	24%

Relationship Goals

...and currently looking for more relationship partners	13	59%
...and currently not looking for more relationship partners	6	27%
Other	3	14%

[[*Bonus points to whoever answered “that’s complicated.” Also, if I’d been thinking I would have included an option for “not looking, but open to the possibility.”*]]

Relationship Status

Single	14	61%
Relationship	6	26%
Married	3	13%
Other	0	0%

Children

0	20	87%
1	1	4%
2	2	9%

More Children

Yes	6	27%
No	8	36%
Uncertain	8	36%

Work and Education

Work Status

Student	14	61%
Academics (on the teaching side)	0	0%
Self-employed	2	9%
Independently wealthy	0	0%
Non-profit work	0	0%
For-profit work	5	22%
Government work	0	0%
Homemaking	1	4%
Unemployed	0	0%
Other	1	4%

Profession

Philosophy	0	0%
Finance / Economics	2	9%
Statistics	0	0%

Computers (AI)	0	0%
Computers (other academic, computer science)	1	4%
Computers (practical: IT, programming, etc.)	8	35%
Physics	1	4%
Biology	3	13%
Medicine	1	4%
Mathematics	2	9%
Neuroscience	2	9%
Other "hard science"	0	0%
Psychology	0	0%
Other "social science"	1	4%
Engineering	2	9%
Business	0	0%
Law	0	0%
Art	0	0%
Other	0	0%

Degree

None	1	5%
High School	10	45%
2 year degree	2	9%
Bachelor's	2	9%
Master's	6	27%
PH D.	1	5%
MD/JD/other professional degree	0	0%
Other	0	0%

Less Wrong

I wanted to get an idea of how our splinter tribe relates to Less Wrong as a whole, so I included enough of a subset of that section from the main survey to do so.

Less Wrong Use

Doesn't use LW	1	5%
Lurker (no account)	6	29%
Lurker (account)	1	5%
Poster (comments only)	5	24%
Poster (discussion)	8	38%
Poster (Main)	0	0%

Time in Community (LW, years)

Mean	1.9
Stddev	1.06
Min	0.2

1st quartile	1
2nd quartile	2
3rd quartile	2.75
Max	4

Karma Score

Mean	183.1
Stddev	265.5
Min	0
1st quartile	0.5
2nd quartile	60
3rd quartile	251
Max	800

Meetups

Yes, regularly.	3	15%
Yes, once or a few times.	6	30%
No.	11	55%

Community (LW)

Yes, all the time.	4	20%
Yes, sometimes.	8	40%
No.	8	40%

Took the main LW survey?

Yes	12	60%
No	8	40%

Less Wrong Study Hall

The meat lies here.

Time in Community (LWSH)

Less than a month.	3	13%
1 - 3 months	4	17%
4 - 6 months	3	13%
6 - 9 months	4	17%
9 - 12 months	9	39%

Frequency

Every day	8	35%
Several times a week	9	39%
Once or twice a week	1	4%
Less than once a week	1	4%
It varies	3	13%

I haven't been here long enough to form a pattern	1	4%

Time in the Hall (per visit, in minutes)

Mean	226.1
Stddev	137.1
Min	60
1st quartile	120
2nd quartile	180
3rd quartile	300
Max	540

[[*NINE HOURS??! O_O I salute you.*]]

Usage

Academic studies	18	27%
Personal projects	21	32%
Deliberate practice (e.g. learning guitar)	4	6%
Work for an employer	5	8%
Chores, paperwork, or other necessities	17	26%
Other	1	2%

[[*I was surprised so many people use it for chores and the like. I do, but I thought I was just about the only one. On the plus side, my home actually gets cleaned up regularly now...*]]

Usage 2 (most important)

Academic studies	12	55%
Personal projects	5	23%
Deliberate practice (e.g. learning guitar)	0	0%
Work for an employer	3	14%
Chores, paperwork, or other necessities	2	9%
Other	0	0%

[[*Unsurprisingly, academics is the big thing.*]]

Draw

Social reinforcement for working.	22	33%
Social punishment for not working.	7	10%
Camera-induced self-consciousness when working.	15	22%
Distraction reduction via group pomodoros.	18	27%
Other	5	7%

Draw 2 (most important)

Social reinforcement for working.	14	61%
Social punishment for not working.	0	0%
Camera-induced self-consciousness when working.	1	4%

Distraction reduction via group pomodoros.	7	30%
Other	1	4%

[[*The carrot is more appreciated than the stick.*]]

Camera

Yes, always.	11	48%
Yes, sometimes.	10	43%
Rarely	1	4%
Never	1	4%

[[*I should have asked about desktop sharing too. I have a suspicion that it works even better than cameras, but only one or two people do it.*]]

Camera Off (why?)

This was a freeform response about why users leave their cameras off. Most users do leave their cameras on so there wasn't a lot to go on here. Included themes were privacy (for oneself or others nearby), temporary non-working states (as when someone is eating), technical difficulties, or just generally not feeling sociable.

Time Zone

UTC+10:00	2	9%
UTC+02:00	1	4.5%
UTC+01:00	13	60%
UTC+00:00	1	4.5%
UTC-05:00	5	23%

[[*Apparently all our American users are on the east coast? Perhaps the West Coast people that might be interested already have access to the main real-life Less Wrong community/CFAR people, and thus don't need the Hall. But I thought we had at least one guy in the West. I am confused.*]]

Temporal Habits (Weekdays)

Mornings (6am-12pm)	6	15%
Afternoons (12pm-5pm)	9	22%
Evenings (5pm-10pm)	18	44%
Late night/very early morning (10pm-6am)	6	15%
Too variable to say	2	5%
I don't use the room during the week	0	0%

Temporal Habits (Weekends)

Mornings (6am-12pm)	4	10%
Afternoons (12pm-5pm)	11	26%
Evenings (5pm-10pm)	13	31%
Late night/very early morning (10pm-6am)	6	14%
Too variable to say	7	17%
I don't use the room on weekends	1	2%

[[*I meant to use time zone information and logon windows to figure out when the room is populated and not, but never got around to it. Maybe next year. Or someone else could do the honors.*]]

Referrals

The initial announcement	9	39%
Other comments or posts on Less Wrong	7	30%
Referred by a friend or partner	6	26%
Referred by a non-LW blog	0	0%
Referred by a search engine	0	0%
Other	1	4%

Interaction (outside the Hall)

Yes, regularly.	9	39%
Yes, sometimes.	7	30%
No.	7	30%

Interaction 2 (in person only)

Yes, regularly.	5	22%
Yes, sometimes.	5	22%
I've met a few people in person once or twice.	2	9%
No.	11	48%

[[*It looks like more people here get involved in-person than Less Wrong as a whole, but it looks like it might be because all our Germans know each other in real life.*]]

Romance

Yes.	5	22%
I didn't meet them through the Hall, but they come there now.	2	9%
No.	16	70%

[[*That Yes% is* much *higher than Less Wrong proper, even though our gender balance isn't much better. Shame our sample size is far too small to tell if it matters. I predict not.*]]

Base Akrasia (1–7 scale, 1 being no akrasia, 7 being functionally paralyzed)

1	1	5%
2	0	0%
3	4	18%
4	5	23%
5	6	27%
6	5	23%
7	1	5%

Akratic Impact (Akrasia in the Hall, same scale as previous)

1	1	5%

2	9	41%
3	8	36%
4	3	14%
5	1	5%
6	0	0%
7	0	0%

[[*Self-reporting being what it is, these numbers aren't terribly reliable. But they're all we've got for the moment and they do show an improvement. My thanks to whoever suggested that two absolute scales were a better idea than absolute + improvement.*]]

Hedonic Impact (Hedonic improvement in the Hall, 1–5, 1 being no/negative impact, 5 being large impact)

1	0	0%
2	0	0%
3	3	13%
4	7	30%
5	13	57%

[[*If I'd been thinking I would have measured hedons in the same manner as akradons. Oh well, next year.*]]

Distractions (percentage)

Mean	29%
Stddev	15
Min	10%
1st quartile	18.5%
2nd quartile	30%
3rd quartile	35%
Max	70%

[[*I am not sure if these numbers are low or high. I would have to compare them to distractibility outside the Hall, but that is much harder to notice and track unless you are explicitly looking for it.*]]

Distraction Cause

People talking or otherwise drawing attention in the Hall during the pomo	0	0%
Digital interruptions (email or IM)	4	17%
In-person interruptions (family or friends wanting attention)	7	30%
Spontaneous web browsing or other computer use.	11	48%
Other	1	4%

[[*I should have done the same checkbox/most-important thing here that I did with the Draw and Usage questions. Unsurprisingly, spontaneous web browsing topped the list by a large margin.*]]

Overwork (percentage)

Mean	34.7
Stddev	24.8
Min	0%
1st quartile	12.5%
2nd quartile	33%
3rd quartile	55%
Max	90%

Accomplishments

Yes	15	65%
No	8	35%

[[*Since we're here for a reason, it seemed good to find out whether people are succeeding at their purpose. Hence this question and the next one.*]]

Accomplishment Examples

This was a freeform question and not everyone answered it. Most answers again involved academic work and study, including one brave soul who studied for and passed final exams for classes on which he'd skipped most of the lectures. I'm not certain if I should applaud that or not. Artistic endeavors were the second most common response, and showed up in the form of painting, manuscripts, fanfiction, and blog posts. One person has been using the Hall to translate HPMoR chapters, and another wrote an open source library. And one of our users landed a new job, for which they'd written their CV using the Hall. Congratulations!

Miscellaneous Feedback

Social Atmosphere

1	9	41%
2	10	45%
3	3	14%
4	0	0%
5	0	0%

[[*Some people have suggested that the recent uptick in sociability and silliness during breaks would distract people. This is an attempt to measure it. More in the personal reflections section.*]]

Suggestion Box

This was a freeform question, and most respondents didn't answer. The most common response by a fair margin was better enforcement of pomodoros, either technologically or socially. There were also multiple calls to replace tinychat.

Despite my complaining, I would be surprised if tinychat was replaced anytime soon. Better social enforcement of pomos should be possible, though. Remember that anyone can track the clock if they want.

Akrasia

Since we're here to fight akrasia, measuring it seemed good. I copied almost the entire Akrasia section from Yvain's survey. A couple of people noted some ambiguous questions. I wanted to keep the data comparable to LW proper, so rather than modifying his questions I added clarified ones in between.

Akrasia

Yes	11	50%
No	11	50%

Akrasia: Current

Yes	2	9%
No	20	91%

Akrasia: Illness

Depression	3	14%
Bipolar disorder	2	10%
Schizophrenia	0	0%
Obsessive-compulsive disorder	0	0%
ADHD	0	0%
PTSD	0	0%
Autism or autism spectrum disorder	2	10%
Generalized anxiety disorder	1	5%
Social phobia	0	0%
Any personality disorder	0	0%
None	12	57%
Other	1	5%

Akrasia: Medicines 1

No	14	67%
Other	7	33%

[[*Someone pointed out that the question as originally stated didn't specify "medications taken specifically to combat akrasia," so I repeated the question with that clarification for both this and the supplements question.*]]

Akrasia: Medicines 1.5

No	16	80%
Other	4	20%

Akrasia: Medicines 2

1	2	25%
2	2	25%
3	3	38%
4	0	0%
5	1	13%

Akrasia: Supplements 1

No	11	55%
Other	9	45%

Akrasia: Supplements 1.5

No	14	74%
Other	5	26%

Akrasia: Supplements 2

1	0	0%
2	1	13%
3	4	50%
4	0	0%
5	3	38%

Akrasia: Therapy 1

No	10	63%
Other	6	38%

Akrasia: Therapy 2

1	1	9%
2	3	27%
3	1	9%
4	1	9%
5	5	45%

Akrasia: Meditation 1

No	9	45%
Other	11	55%

Akrasia: Meditation 2

1	2	17%
2	4	33%
3	4	33%
4	1	8%
5	1	8%

Akrasia: Elsewhat 1

No	9	50%
Other	9	50%

Akrasia: Elsewhat 2

1	3	27%
2	3	27%
3	3	27%
4	1	9%

5	1	9%

Akrasia: Communication

Yes	16	76%
No	5	24%

Silliness

Some things suggested by our users when I was putting the survey together. This is just for fun.

Tinychat Hatred

1	4	18% (thousand burning suns)
2	1	5%
3	7	32%
4	3	14%
5	7	32% (Emperor Palpatine)

Tinychat Screams (number of)

I don't know how to correctly calculate averages or related statistics when some numbers are written in Knuth notation, so there's not much to report here. My favorite answer was Pi^^^42.

Stuffies

No	5	24%
Yes, but only one	2	10%
Yes, more than one	13	62%
Tons	1	5%

Stuffies on Camera

Yes	10	48%
No	11	52%

[[*For some reason lots of us have stuffies.*]]

And that's the end of it. My thanks to Lachouette for feedback on a draft of this post, and to the Hall itself for keeping me on track while I was writing about it. Those who want to join us, consider yourselves invited.

---

1. I'm posting this to Main, but given that our splinter tribe is a very small fragment of Less Wrong, I'm not sure if it's appropriate to Main. If a mod wants to move it, I won't object. ↩

2. Someday we'll have users with bizarre half- or quarter-hour timezone offsets and this will all fall down. ↩

3. While there's no explicit rules about discussion topics, it's probably a good idea to observe Less Wrong's anti-politics norm. The few times I've witnessed

conflicts between regulars, it was when politics came up in some way. ↩

Don't teach people how to reach the top of a hill

When is it faster to rediscover something on your own than to learn it from someone who already knows it?

Sometimes it's faster to re-derive a proof or algorithm than to look it up. Keith Lynch re-invented the fast Fourier transform because he was too lazy to walk all the way to the library to get a book on it, although that's an extreme example. But if you have a complicated proof already laid out before you, and you are not Marc Drexler, it's generally faster to read it than to derive a new one. Yet I found a knowledge-intensive task where it would have been much faster to tell someone nothing at all than to tell them how to do it.

I'm digitizing my books by chopping off their bindings and scanning them. I recently hired someone to do the chopping, and have been teaching him how to do it. The first step is to chop the book into sections of about 50 pages each, separating them at the binding. I do this by placing the opened book cover-down under a paper chopper, and cutting it precisely where the two opened pages meet.

The "chopper" is a manual paper-cutter with a 15-inch steel blade that weights about 10 pounds and is razor-sharp. If the blade is a fifth of a millimeter off its mark, it misses the gap between the pages and makes the cutting much harder, as it must go through paper instead of only glue. Being an entire millimeter off makes the blade catch the page maybe half a centimeter further away from its edge, depending on how the base of the page is angled, cutting off words and ruining the book. You can't see where the blade touches the book while making the cut. You can look before making the cut and position the book, but then you need one hand to operate the blade, and the physics of a book that wishes to spring shut, fall away from the blade, and fall onto one side, make it nearly impossible to keep the groove in the book in place for the blade with just one hand, unless you hold it with your fingers underneath the blade, which you can do only once.

My technique is to do this:

1. Face the chopper with the blade's hinge opposite you. The chopper has a square base. If you're facing north, the hinge is at its northeast corner; the blade will cut along the eastern edge.
2. Slide the book, cover-down, under the blade until the blade is over the binding.
3. Lower the blade until it almost touches the book.
4. Grasp the book with 2 hands, one on each side of the blade. Lift the book in the air to touch the blade along its entire length. (The edge of the blade where the cut begins is lower than the other edge.)
5. Slide the book back and forth until the groove between the pages locks into place against the blade.
6. Lower the blade until the end that starts the cut is pressed firmly against the book, which is pressed firmly against the cutting board.
7. Press your left fingers against the pages to be cut off from the book, pressing those pages up against the blade along its entire length and keeping the groove of the book in place along the blade. You can feel the side of the blade through the pages, but the blade is now too low for your fingers to get underneath it.

8. Stand with your head and shoulders directly above the blade. DO NOT raise the blade while repositioning yourself.
9. Punch downward with the blade while simultaneously falling on it with all your weight to make a cut that is too fast to grab onto the paper and pull it out of place.

It's more complicated than that; I'm simplifying for the sake of space. But I didn't realize any of this when I began teaching him. I told him to put the book face-up under the blade and cut it into sections. It takes me a few seconds to make each cut. It was only when he kept trying to do it, and it kept not working, that I realized there must be more to it. He would try to chop a book and it wouldn't work. I'd look at the book, figure out what went wrong, then chop another section from the same book, watching myself, until I figured out what I was doing that could make the difference. Then I'd tell him. He'd try again, and it still wouldn't work. After perhaps 3 hours (6 person-hours), we worked out the sequence of steps I was doing well enough that he could chop books.

I could ask how I learned all those steps without knowing I'd learned them--was I conscious of them at the time, but forgot each step as soon as it was committed to my body? Probably. And it's interesting that I was unable to extract my own knowledge without watching someone else fail. But my point in this essay is that it took me longer to teach him how to do it than it took me to learn how to do it on my own--and it took 2 people instead of 1. So teaching was less than half as efficient as just handing him a book and walking away. (I'm ignoring the risk of coming back an hour later to find the floor strewn with severed fingers; that's overly particular to this domain.)

He kept worrying whether he was "doing it right". When I first figured out how to use the book chopper, I didn't know if there was one right way, or five, or none. I didn't have anyone to compare myself to. I could see whether I'd chopped the book the way I wanted to, but had no way to judge whether I was doing "above average", and so no self-consciousness about how well I was doing. Whereas he would see me take a book, slide it in, and chop it correctly, and then he would spend minutes fiddling with it, bending down to look under the blade from each side, swapping left and right hands between the two sides of the book and the blade, raising and lowering the blade, ad nauseum, until he finally tried to cut it--and inevitably got it wrong. He was nearly disabled by frustration and a sense of incompetence, and his actions were the anxious, tentative movements of someone worried about "doing it wrong" rather than the rapid movements of someone trying to find out whether there was any way to do it at all. We often hear the inspirational advice that believing something is possible makes it easier to accomplish; yet I saw just the opposite here. I didn't have my self-image on the line in my initial discovery process because I didn't know whether my task was possible, so I felt no pressure.

The task I originally faced was to find any path through a very large space that would end up with a book cut where I wanted it cut. The task the two of us faced in teaching him was to observe me chopping books, over and over, until we could find the one path I had discovered and forgotten. It isn't obvious which of these tasks is easier. In "discovery", there may be many possible solutions, while in "imitation" there is only one.

The psychological component probably applies to every search space: Availability of experts and the belief that there is a right way to do something inhibit experimentation; focusing on imitation prevents discovery. But what was it about the

search space for book-chopping that made experimentation simple enough, and imitation hard enough, that imitation was harder than discovery? My guess is it was these things:

- The task is analog/continuous, concerned with movements in space, so that it can't be specified precisely.
- The task is procedural, and almost all of the teacher's knowledge about it is "motor memory", not conscious.
- The search space is low-dimensional, because except for the final act of cutting and the importance of keeping fingers out of the path of the blade, every action involves only the book and the paper chopper. The book and the chopper each have one degree of freedom, and the book can be moved in space, and that is all.
- There are no difficult insights, special sticking-points, or especially-valuable insights that could be applied repeatedly (discontinuities in the search space).
- There was little back-tracking in rediscovering the steps. So there are few local maxima.
- Failures are easy to analyze; the next step in the process can be discovered by analyzing the previous failure. Also, steps 3-5, 6-7, and 8-9 are mostly independent; e.g., you can discover steps 8-9 before having 6-7 completely worked out. These properties allow the task to be learned incrementally.

Roughly, it's a task in a search space on which hill-climbing works well.

Contrast this to martial arts, in which the movements of two fighters have a much higher dimensionality. The fraction of all possible movement sequences that leads to a side-kick or a hook punch is so small that few boxers ever discover the first and few karate students ever discover the second. Or contrast it to mathematical proofs, which are very high-dimensional, may have key insights (discontinuities), and give little indication of whether one is making progress. Those are domains in which instruction is more useful.

Think about computer software that you had to read the manual for. I think first of Adobe Photoshop and its concept of layers and selections. Those are complex, broadly-applicable concepts (discontinuities) that you can't easily discover by experimentation, as clicking on things before you understand them will make apparently random things happen. A user interface for something casual (a game, a website) or meant for the mass-market should have an event space on which hill-climbing works well, so that instruction is not needed.

On not getting a job as an option

This was originally a comment to VipulNaik's recent indagations about the academic lifestyle versus the job lifestyle. Instead of calling it lifestyle he called them career options, but I'm taking a different emphasis here on purpose.

Due to information hazards risks, I recommend that Effective Altruists who are still wavering back and forth do not read this. Spoiler EA alert.

I'd just like to provide a cultural difference information that I have consistently noted between Americans and Brazilians which seems relevant here.

To have a job and work in the US is taken as a *de facto* biological need. It is as abnormal for an American, in my experience, to consider not working, as it is to consider not breathing, or not eating. It just doesn't cross people's minds.

If anyone has insight above and beyond "Protestant ethics and the spirit of capitalism" let me know about it, I've been waiting for the "why?" for years.

So yeah, let me remind people that you can spend years and years not working. that not getting a job isn't going to kill you or make you less healthy, that ultravagabonding is possible and feasible and many do it for over six months a year, that I have a friend who lives as the boyfriend of his sponsor's wife in a triad and somehow never worked a day in his life (the husband of the triad pays it all, both men are straight). That I've hosted an Argentinian who left graduate economics for two years to randomly travel the world, ended up in Rome and passed by here in his way back, through couchsurfing. That Puneet Sahani has been well over two years travelling the world with no money and an Indian passport now. I've also hosted a lovely estonian gentleman who works on computers 4 months a year in London to earn pounds, and spends eight months a year getting to know countries while learning their culture etc... Brazil was his third country.

Oh, and never forget the Uruguay couple I just met at a dance festival who have been travelling as hippies around and around South America for 5 years now, and showed no sign of owning more than 500 dollars worth of stuff.

Also in case you'd like to live in a paradise valley taking Santo Daime (a religious ritual with DMT) about twice a week, you can do it with a salary of aproximatelly 500 dollars per month in Vale do Gamarra, where I just spent carnival, that is what the guy who drove us back did. Given Brazilian or Turkish returns on investment, that would cost you 50 000 bucks in case you refused to work within the land itself for the 500.

Oh, I forgot to mention that though it certainly makes you unable to do expensive stuff, thus removing the paradox of choice and part of your existential angst from you (uhuu less choices!), there is nearly no detraction in status from not having a job. In fact, during these years in which I was either being an EA and directing an NGO, or studying on my own, or doing a Masters (which, let's agree is not very time consuming) my status has increased steadily, and many opportunities would have been lost if I had a job that wouldn't let me move freely. Things like being invited as Visiting Scholar to Singularity Institute, like giving a TED talk, like directing IERFH, and

like spending a month working at FHI with Bostrom, Sandberg, and the classic Lesswrong poster Stuart Armstrong.

So when thinking about what to do with you future my dear fellow Americans, please, at least consider not getting a job. At least admit what everyone knows from the bottom of their hearts, that jobs are abundant for high IQ people (specially you my programmer lurker readers.... I know you are there...and you native English speakers, I can see you there, unnecessarily worrying about your earning potential).

A job is truly an instrumental goal, and your terminal goals certainly do have chains of causation leading to them that do not contain a job for 330 days a year. Unless you are a workaholic who experiences flow in virtue of pursuing instrumental goals. Then please, work all day long, donate as much as you can, and may your life be awesome!