

# Best of LessWrong: February 2016

1. [Require contributions in advance](#)
2. [Unofficial Canon on Applied Rationality](#)
3. [The Brain Preservation Foundation's Small Mammalian Brain Prize won](#)
4. [Defiance](#)
5. [Conclusion of the Replacing Guilt series](#)
6. [Recklessness](#)
7. [How we will be measured](#)

## Best of LessWrong: February 2016

1. [Require contributions in advance](#)
2. [Unofficial Canon on Applied Rationality](#)
3. [The Brain Preservation Foundation's Small Mammalian Brain Prize won](#)
4. [Defiance](#)
5. [Conclusion of the Replacing Guilt series](#)
6. [Recklessness](#)
7. [How we will be measured](#)

# Require contributions in advance

If you are a person who finds it difficult to tell "no" to their friends, this one weird trick may save you a lot of time!

## Scenario 1

Alice: "Hi Bob! You are a programmer, right?"

Bob: "Hi Alice! Yes, I am."

Alice: "I have this cool idea, but I need someone to help me. I am not good with computers, and I need someone smart whom I could trust, so they wouldn't steal my idea. Would you have a moment to listen to me?"

Alice explains to Bob her idea that would completely change the world. Well, at the least the world of bicycle shopping.

Instead of having many shops for bicycles, there could be one huge e-shop that would collect *all* the information about bicycles from *all* the existing shops. The customers would specify what kind of a bike they want (and where they live), and the system would find all bikes that fit the specification, and display them ordered by lowest price, including the price of delivery; then it would redirect them to the specific page of the specific vendor. Customers would love to use this one website, instead of having to visit multiple shops and compare. And the vendors would have to use this shop, because that's where the customers would be. Taking a fraction of a percent from the sales could make Alice (and also Bob, if he helps her) incredibly rich.

Bob is skeptical about it. The project suffers from the obvious chicken-and-egg problem: without vendors already there, the customers will not come (and if they come by accident, they will quickly leave, never to return again); and without customers already there, there is no reason for the vendors to cooperate. There are a few ways how to approach this problem, but the fact that Alice didn't even think about it is a red flag. She also has no idea who are the big players in the world of bicycle selling; and generally she didn't do her homework. But after pointing out all these objections, Alice still remains super enthusiastic about the project. She promises she will take care about everything -- she just cannot write code, and she needs Bob's help for this part.

Bob believes strongly in the division of labor, and that friends should help each other. He considers Alice his friend, and he will likely need some help from her in the future. Fact is, with perfect specification, he *could* make the webpage in a week or two. But he considers bicycles to be an extremely boring topic, so he wants to spend as little time as possible on this project. Finally, he has an idea:

"Okay, Alice, I *will* make the website for you. But *first* I need to know exactly how the page will look like, so that I don't have to keep changing it over and over again. So here is the homework for you -- take a pen and paper, and make a sketch of how exactly the web will look like. All the dialogs, all the buttons. Don't forget logging in and logging out, editing the customer profile, and everything else that is necessary for the website to work as intended. Just look at the papers and imagine that you are the

customer: where exactly would you click to register, and to find the bicycle you want? Same for the vendor. And possibly a site administrator. Also give me the list of criteria people will use to find the bike they want. Size, weight, color, radius of wheels, what else? And when you have it all ready, I will make the first version of the website. But until then, I am not writing any code."

Alice leaves, satisfied with the outcome.

This happened a year ago.

No, Alice doesn't have the design ready, yet. Once in a while, when she meets Bob, she smiles at him and apologizes that she didn't have the time to start working on the design. Bob smiles back and says it's okay, he'll wait. Then they change the topic.

## **Scenario 2**

Cyril: "Hi Diana! You speak Spanish, right?"

Diana: "Hi Cyril! Yes, I do."

Cyril: "You know, I think Spanish is the most cool language ever, and I would really love to learn it! Could you please give me some Spanish lessons, once in a while? I totally want to become fluent in Spanish, so I could travel to Spanish-speaking countries and experience their culture and food. Would you please help me?"

Diana is happy that someone takes interest in her favorite hobby. It would be nice to have someone around she could practice Spanish conversation with. The first instinct is to say yes.

But then she remembers (she knows Cyril for some time; they have a lot of friends in common, so they meet quite regularly) that Cyril is always super enthusiastic about something he is totally going to do... but when she meets him next time, he is super enthusiastic about something completely different; and she never heard about him doing anything serious about his previous dreams.

Also, Cyril seems to seriously underestimate how much time does it take to learn a foreign language fluently. Some lessons, once in a while will not do it. He also needs to study on his own. Preferably every day, but twice a week is probably a minimum, if he hopes to speak the language fluently within a year. Diana would be happy to teach someone Spanish, but not if her effort will most likely be wasted.

Diana: "Cyril, there is this great website called Duolingo, where you can learn Spanish online completely free. If you give it about ten minutes every day, maybe after a few months you will be able to speak fluently. And anytime we meet, we can practice the vocabulary you have already learned."

This would be the best option for Diana. No work, and another opportunity to practice. But Cyril insists:

"It's not the same without the live teacher. When I read something from the textbook, I cannot ask additional questions. The words that are taught are often unrelated to the

topics I am interested in. I am afraid I will just get stuck with the... whatever was the website that you mentioned."

For Diana this feels like a red flag. Sure, textbooks are *not* optimal. They contain many words that the student will not use frequently, and will soon forget them. On the other hand, the grammar is always useful; and Diana doesn't want to waste her time explaining the basic grammar that any textbook could explain instead. If Cyril learns the grammar and some basic vocabulary, *then* she can teach him all the specialized vocabulary he is interested in. But now it feels like Cyril wants to avoid *all* work. She has to draw a line:

"Cyril, this is the address of the website." She takes his notebook and writes 'www.duolingo.com'. "You register there, choose Spanish, and click on the first lesson. It is interactive, and it will not take you more than ten minutes. *If* you get stuck there, write here what exactly it was that you didn't understand; I will explain it when we meet. If there is no problem, continue with the second lesson, and so on. When we meet next time, tell me which lessons you have completed, and we will talk about them. Okay?"

Cyril nods reluctantly.

This happened a year ago.

Cyril and Diana have met repeatedly during the year, but Cyril never brought up the topic of Spanish language again.

### **Scenario 3**

Erika: "Filip, would you give me a massage?"

Filip: "Yeah, sure. The lotion is in the next room; bring it to me!"

Erika brings the massage lotion and lies on the bed. Filip massages her back. Then they make out and have sex.

This happened a year ago. Erika and Filip are still a happy couple.

Filip's previous relationships didn't work well, in long term. In retrospect, they all followed a similar scenario. At the beginning, everything seemed great. Then at some moment the girl started acting... *unreasonably*?... asking Filip to do various things for her, and then acting annoyed when Filip did exactly what he was asked to do. This happened more and more frequently, and at some moment she broke up with him. Sometimes she provided explanation for breaking up that Filip was unable to decipher.

Filip has a friend who is a successful salesman. Successful both professionally and with women. When Filip admitted to himself that he is unable to solve the problem on his own, he asked his friend for advice.

"It's because you're a f\*\*\*ing doormat," said the friend. "The moment a woman asks you to do anything, you immediately jump and do it, like a well-trained puppy. Puppies

are *cute*, but not *attractive*. Have you ready any of those books I sent you, like, ten years ago? I bet you didn't. Well, it's all there."

Filip sighed: "Look, I'm not trying to become a pick-up artist. Or a salesman. Or anything. No offense, but I'm not like you, personality-wise, I never have been, and I don't want to become your - or anyone else's - copy. Even if it would mean greater success in anything. I prefer to treat other people just like I would want them to treat me. Most people reciprocate nice behavior; and those who don't, well, I avoid them as much as possible. This works well with my friends. It also works with the girls... at the beginning... but then somehow... uhm... Anyway, all your books are about manipulating people, which is ethically unacceptable for me. Isn't there some other way?"

"All human interaction *is* manipulation; the choice is between doing it right or wrong, acting consciously or driven by your old habits..." started the friend, but then he gave up. "Okay, I see you're not interested. Just let me show you the most obvious mistake you make. You believe that when you are nice to people, they will perceive you as nice, and most of them will reciprocate. And when you act like an asshole, it's the other way round. That's correct, on some level; and in a perfect world this would be the whole truth. But on a different level, people also perceive nice behavior as *weakness*; especially if you do it habitually, as if you don't have any other option. And being an asshole obviously signals *strength*: you are not afraid to make other people angry. Also, in long term, people become used to your behavior, good or bad. The nice people don't seem so nice anymore, but they still seem weak. Then, ironically, if the person well-known to be nice refuses to do something *once*, people become really angry, because their expectations were violated. And if the asshole decides to do something nice *once*, they will praise him, because he surprised them pleasantly. You should be an asshole once in a while, to make people see that you have a choice, so they won't take your niceness for granted. Or if your girlfriend wants something from you, sometimes just say no, even if you *could* have done it. She will respect you more, and then she will enjoy more the things you *do* for her."

Filip: "Well, I... probably couldn't do that. I mean, what you say seems to make sense, however much I hate to admit it. But I can't imagine doing it myself, especially to a person I love. It's just... uhm... *wrong*."

"Then, I guess, the very least you could do is to ask *her* to do something for *you* first. Even if it's symbolic, that doesn't matter; human relationships are mostly about role-playing anyway. Don't jump immediately when you are told to; always make *her* jump first, if only a little. That will demonstrate strength without hurting anyone. Could you do that?"

Filip wasn't sure, but at the next opportunity he tried it, and it worked. And it kept working. Maybe it was all just a coincidence, maybe it was a placebo effect, but Filip doesn't mind. At first it felt kinda artificial, but then it became natural. And later, to his surprise, Filip realized that practicing these symbolic demands actually makes it easier to ask when he really *needed* something. (In which case sometimes *he* was asked to do something first, because his girlfriend -- knowingly or not? he never had the courage to ask -- copied the pattern; or maybe she has already known it long before. But he didn't mind that either.)

The lesson is: If you find yourself repeatedly in situations where people ask you to do something for them, but at the end they don't seem to appreciate what you did for them, or don't even care about the thing they asked you to do... and yet you find it difficult to say "no"... ask *them* to contribute to the project *first*.

This will help you get rid of the projects they don't care about (including the ones they *think* they care about in *far* mode, but do not care about enough to actually work on them in *near* mode) without being the one who refuses cooperation. Also, the act of asking the other person to contribute, after being asked to do something for them, mitigates the status loss inherent in working for them.

# Unofficial Canon on Applied Rationality

I have been thinking for a while that it would be useful if there was something similar to the [Less Wrong Canon on Rationality](#) for the CFAR material. Maybe, it could be called the 'CFAR Canon on Applied Rationality'. To start on this I have compiled a collection of descriptions for the CFAR techniques that I could find. I have separated the techniques into a few different sections. The sections and descriptions have mostly been written by me, with a lot of borrowing from other material, which means that they may not accurately reflect what CFAR actually teaches.

Please note that I have not attended any [CFAR workshops](#), nor am I affiliated with [CFAR](#) in any way. My understanding of these techniques comes from CFAR videos, blogs and other websites which I have provided links to. If I have missed any important techniques or if my understanding of any of the techniques is incorrect or if you can provide links to the research that these techniques are based on, please let me know and I will update this post.

## Warning:

Learning this material based solely on the descriptions written here may be unhelpful, arduous or even harmful. (See Duncan\_Sabien's [full comment](#) for more information on this) It is because the material is very hard to learn correctly. Most of the techniques below involve in one way or another volitionally overriding your instinctual, intuitive or ingrained behaviours and thoughts. These are thoughts which not only often feel enticing and alluring, but that also often feel unmistakably right. If you are anything like me, then you should be very careful if you are trying to learn this material alone. For you will be prone to rationalization, taking shortcuts and making mistakes.

My recommendations for trying to learn this material are:

- learn it deeply and be sure to put what you have learnt into practice. It will often help if you take notes on what works for you and what doesn't. Also take note of the 'Mindsets and perspectives that help you in discovering potential situations that you could end up valuing' section as these are very important.
- get the help of experts or other people who have already expended great amounts of effort in trying to implement this material like the people at cfar. This will save you a great amount of stress and effort as it will allow you to avoid a plethora of potential mistakes and inefficiencies. If you really want to learn this material, then you should deeply consider attending a CFAR workshop.
- get the help of or involve friends. As Duncan\_Sabien has [said](#):

It is better on almost every axis with instructors, mentors, friends, companions—people to help you avoid the biggest pitfalls, help you understand the subtle points, tease apart the interesting implications, shore up your motivation, assist you in seeing your own mistakes and weaknesses. None of that is impossible on your own, but it's somewhere between one and two orders of magnitude more efficient and more efficacious with guidance".

- be dubious of your mental models. Beware thoughts and ideas that feel unequivocally right especially if they are solely located internally rather than also being expressed or formulated externally.



- You might want to bookmark this page instead of reading it all at once as it is quite long.

## Sections:

- [Bugs](#)
  - [Discovering bugs](#)
  - [Things that are probably bugs and should be analysed](#)
  - [Analyzing bugs](#)
  - [Solving bugs](#)
  - [Mindsets and perspectives that help you in discovering potential situations that you could end up valuing](#)
  - [Optimizing your ability to follow through with your plans](#)
  - [Learning optimization](#)
  - [Planning optimization](#)
  - [Improving the accuracy of your credence levels](#)
  - [Mindfulness of how your internal state is affecting how you perceive the world](#)
  - [Resolving Disagreements](#)
  - [Other stuff which has already been covered in LessWrong](#)
- 

## Bugs

An important concept that is required to understand the CFAR material is the concept of '[bugs](#)'. Bugs generally tend to be situations that involve a feeling of "stuckness" and often occur when your system one and two wants are out of alignment. Some concrete examples of bugs include:

- things that aren't working and you don't know why, e.g. repeatedly not getting things done on time or always having issues with your finished products at work
- things that just don't feel right, e.g. your daily schedule or personal relationships
- things that you want to do, but also don't want to do, e.g. exercise or doing your taxes
- things you want to improve and don't know how to, e.g. getting a good night's sleep or social skills
- things that you want to do, but are afraid to do, e.g. public speaking
- plans you think are going to fail, e.g. getting your class assignment in on time
- other dissatisfactions and inefficiencies
- various other internal conflicts

CFAR would stress that 'bugs' are not things that should be accepted with resignation. They are instead things that should be worked through and solved. They are problems that deserve your time, attention and courage to solve. Due to our human nature, it is often best to get the help [of others](#) when you are trying to solve your bugs as we tend to rationalize and justify our bugs.

In summary, the CFAR perspective on bugs seems to be that when you notice one you should think: "Okay! Here is an opportunity for me to get better at life. Where's my pen and paper?" or "Where can I find someone to talk this through with".

**Discovering bugs** - the below techniques all deal, in one way or another, with improving your ability to be able to discover your bugs.

- [Hamming questions](#) - the mathematician Richard Hamming was known to approach experts from other fields and ask "what are the important problems in

your field, and why aren't you working on them?". The same question can be applied to personal life: "what are the important problems in your life and what is stopping you from working on them?" This question is often best asked in what for convenience is often called a "[hamming circle](#)". This is a group of people who come together and help each other explore their hamming questions. As a word of caution, [jimrandomh](#) has advised that:

If you organize a group into Hamming Circles and they don't know what they're doing, aren't in the right mindspace, or don't have enough shared context and trust, it can backfire pretty severely. People's Hamming problems are often things that are aversive to think about, and attempting to discuss them but having it go poorly can make the problem worse.

- [The Surprise Journal](#) – is a simple technique which involves recording when and why you are surprised. Surprise is a cue that your expectations are and were wrong. Part of the usefulness of this technique may be that it primes the reticular activation system so that you are on the lookout for and notice the things that surprise you.
- [Comfort Zone Expansion \(CoZE\)](#) - is basically CFAR's take on [exposure therapy](#). Now, most people don't have full on debilitating disorders, but it is common for people to accumulate some negative, avoidant or escape-oriented strategies. Some example situations in which these strategies tend to occur can be found [here](#). These strategies tend to form for a large number of reasons, but the most common is probably due to our human brains being bad at evaluating the negative social impacts and consequences of our actions. This is due to the [negativity bias](#), as well as other things. CoZE is a way to fix these strategies. It could be described as a long-term method to better calibrate your evaluations or also as applied aversion factoring. There are two main goals involved in it. The first is to train yourself to be able to do things that you find aversive, i.e. to be able to feel the fear and do it anyway. The second is to remove or lessen your non helpful aversions. These two goals are entwined as the idea of CoZE is to frequently expose you to the aversive thing (in a safe way) and to make sure you do it despite the discomfort. The idea is that this will help you to realize that it is not as bad as you predict. Some tips if you are going to try CoZE are: find some way to involve others either as an audience or as a way to provide accountability, care simply about getting what you want to do done rather than getting it done well or with good results and try to cultivate a culture or sense of [playfulness](#) as this leads to better coping styles for dealing with stress. CoZE involves:
  - Creating a list of doable, but also challenging situations that are outside your comfort zone, but not by so much that they are going to cause trauma or any other long term negative effects.
  - Noting down the difficulty of the task in accordance with your comfort zone. The difficulty of a task is dependent on the individual. A lawyer might have no difficulties with social situations, but could find situations involving heights to be extremely difficult. Another person might be the exact opposite.
  - Doing the aversive thing. It is good idea to view the discomfort as a part of the growth experience. If it is uncomfortable, then it means that you are extending yourself and what you are capable of. The discomfort is the price you pay for advancing towards a goal.
  - Note down or think about whether the real consequences matched up with your predictions. Even if you do something that turns out to genuinely dangerous or harmful, then this is good (as long as you are fine) as it has

provided you with some reasonably trustworthy information that you should avoid the thing you just did or at least adapt your strategies with engaging with it.

- [Directed graphical causal models for our personal lives](#) - this [modelling process](#), which probably happens more naturally for some of us than others, can help us focus locally on relevant causal relationships, and generate queries about them that we might otherwise forget to seriously ask.
- [Value of information and fermi-type estimation in daily life](#) - an example of this is asking: "how many minutes will I spend commuting over the next year and what's the expected savings if I set a 5-minute timer to try to optimize that?" The basic idea is that you are trying to find valuable information that would allow you to estimate or improve the value of some large and expensive (or frequent) task that you plan to do in the future. For example, many people commit to years of education and tens of thousands of hours of work to "get a good job doing X" without even spending as little as 40 hours job-shadowing to see if they'd actually enjoy doing X, or 10 hours cleverly searching for alternative lines of work.

**Things that are probably bugs and should be analysed** - the below are not really techniques, but are instead descriptions of particular situations that frequently turn out to be bugs.

- [Alienated birth rights](#) - these are things that all human beings should be capable of engaging in and enjoying, but which people often do not because of internalized beliefs and identities that say things like: "I cannot do that" or "I am bad at that". (This is similar to [learned blankness](#) and [learned helplessness](#))
- Half-hearted trying often induces failure and is rarely the optimal behaviour (this is my interpretation of the material in [this post](#)) - 'half-hearted trying' could be described as either handicapping or self-limiting behaviour as well as when you are just going through the motions rather than fully applying yourself. Half-hearted trying is a common response when someone is apathetic or predicting that failure is inevitable, or at least highly probable. It is different from cautious, tentative and exploratory behaviours. Half-hearted trying isn't necessarily a bad thing. It is often totally fine to be doing it and you shouldn't view it as anything more than an indication that you don't really care about what you are doing. You shouldn't feel guilty when you notice that you are half-heartedly trying. In fact, finding this out should be a cause for celebration as it means that you have just found a new reliable source of information into your true motivations. This new information can then be used to debug why you are half-heartedly trying. Successful debugging will generally lead to four types of situations. If you find out that what you are half-heartedly trying to do is something that you shouldn't care about, then you should spend some time and effort to find out if it's possible to stop doing it. If you find out that it is something that you should really be caring about more, then you have found a bug that you should solve. If you find out that you are handicapping yourself because you fear failure, then you should try to remove these fears or aversions, see the 'Taking apart your aversions' section for more on this. Finally, if you find out that it is something that you need to do, but shouldn't care about, then you should try to accept your half-hearted trying. You should not punish yourself for half-heartedly trying because, at the end of the day, even if you are doing something half-heartedly at least you are doing it. As [kurt Vonnegut](#) once wrote: "If you can do a half-assed job of anything, you're a one-eyed man in a kingdom of the blind". Perhaps, we should only concern ourselves over those tasks that we repeatedly find ourselves half-heartedly trying to do. We should beware of the fact that

callousness and apathy often come easily. Consistently caring, on the other hand, can require great amounts of character and courage. Some examples of half-heartedly trying include when:

- Your goal is just to get something done rather than to get it done well. For example, someone might be nagging you to do something and so you reluctantly and half-heartedly do it.
- You find something to be pointless, but are unable to stop doing it. For example, you might be in a relationship that you no longer want to be in, but you don't have the courage to break it up, so you withdraw physically/emotionally and continue the relationship, but are only half-committed to it.
- You find that it makes failure more palatable or less socially costly. For example, if you do end up failing you like the idea that you can always say: "Oh, yeh, but I wasn't really trying"
- You highly value giving up early in certain contexts. For example, [beginner parkour students often do not do certain moves well](#) because it is easy to quit early in them.
- You find some aspect of the activity to be aversive, but you still want to do it. For example, you might feel guilty about doing something, so you end up only half committing to doing it.
- Remove wasteful and/or harmful repeated thoughts (this is my interpretation of the material in [this post](#)) - rumination and dwelling is a very human way of solving complex problems, but ruminations are expensive and can also alter us in significant ways. It is therefore important that you make sure that your ruminations are directed at worthy and solvable problems. Some different types of problems that we ruminate about include:
  - Problems that System 1 can solve by itself : these are good leave them alone
  - Problems that are worth solving, but require System 2 input : this would include repeated behaviours that you dislike. For example, feeling stuck in a job or relationship, but doing nothing about it. The solution here is to raise the problem to conscious attention and then try to figure out what is bothering system 1. You should then decide on what to do about it.
  - "Problems" that should be accepted : consider someone who is making an annoying munching sound. Most people would think in their heads: "will you shutup" or something similar. These negative thoughts do nothing to change the actual situation. The best thing to do in this situation is to decide on a strategy to either stop the aggravation or to accept it and stop dwelling negatively on it.
  - Problems that should be delegated either to others or to future you : simply put, if dwelling on something is not helpful in anyway then it should be stopped. For example, if on the way home after an interview you are thinking about all the things that you might have messed up then you are effectively causing yourself stress for no reason. This is because there is no way for you to take actions on these thoughts. You cannot retake the interview and so you shouldn't worry about what might have gone wrong. It is ok to take notes on areas to improve. It is a problem, however, when you dwell on these areas outside of situations where you can take actions to improve in them. In the same way if you are dwelling on something that others must do or that you can only do in the future then these thoughts are effectively wasteful.
  - Problems that require dwelling or rumination to solve : examples include complex academic problems that you are trying to solve. Rumination and dwelling in this case is useful and you should let it be.

**Analyzing bugs** - these techniques are all about helping you to better understand what you actually value or find aversive

- [Goal factoring](#) - this entails finding some alternate set of actions through which you could get what you want cheaper. For instance, if you believe that you are doing martial arts solely for the exercise and self-defense benefits, but do not want to get these things by, for example, jogging to work and carrying mace, then it indicates that there is something more to your positive evaluation of martial arts. You might, for example, want to do it because you think it's cool or because your friends do it. By listening to your emotional reactions and analyzing them you can find out how to better align your evaluations so that you end up correctly valuing what you actually find valuable.
- [Reference class hopping](#) - while [reference class forecasting](#) is about improving your reference class so that you can make more accurate predictions. Reference class hopping is about discovering the underlying cause of anxiety or [akrasia](#). It does this by exploring different reference classes and tests them to find out if they arouse anxiety or akrasia. For example, imagine that you dislike social gatherings. Reference class hopping would tell you to look at your track record to consider whether you feel anxious at all parties or whether there are times when the anxiety is diminished. By identifying the common themes in the parties that you find uncomfortable, you can develop a hypothesis about the source of your anxiety and then test the theory.

## Solving bugs

- [Taking apart your aversions](#) or [aversion factoring](#) - involves:
  - Finding a task that you find aversive, but think might be worthwhile doing
  - Attempting to list the complete set of aversions, i.e. everything that is stopping you from doing the task
  - Making each aversion concrete, i.e. finding out what exactly, at the smallest level possible, is aversive. We often think of aversions in the conglomerate sense. We think "I don't like social gatherings" instead of "I don't like aspect X of social gatherings". The problem with this is that it leads us to find things in their entirety to be aversive when it might really only be one aspect of it that we find to be aversive.
  - Solving your aversions. The strategy that you take here will depend on the type of aversion that it is. We can think of aversions as being one of two types: helpful and non helpful.
    - A helpful aversion is one where you would think: "Oh, thanks brain for helping me avoid that harmful situation". Aversions of this type should either be accepted or solved by generating procedures or techniques that eradicate the danger or potentially harmful situation. Once the aversion is solved it may still linger, which means it has turned into a non helpful aversion.
    - A non helpful aversion is one that is present when there is no danger or reason for its existence. Aversions tend to be sticky and residual which means that people's aversions often prompted helpful behaviors in the past, but since their environment has changed they have now become unhelpful. The goal in this step is to simply develop a plan that will help to remove these aversions. Often this will involve acclimation or exposure. Reframing can also help. Also, see the section: "optimising your ability to follow through with your plans"

- Doing a mindful walk through of the activity that was originally aversive to see if it is still aversive. If it is, then repeat this process from step 2.
- [Reframing](#) - when making decisions the underlying valuation of a solution or problem depends on the perspective or way in which you are framing the solution or problem. For example, instead of thinking "Do I want to go the gym?" you can think "If I was already at the gym, would I want to leave and wish that I hadn't gone?"
- [Pitting your desire to look good against your desire to improve](#) - if you are going to improve fast, then you will probably also need to fail fast and fail frequently, at least at first. You will need to be doing things like: asking 'silly' questions or trying stuff out even when you're not ready. This can be hard to do because no one likes to look like a fool. One technique that sometimes helps in making this easier is to compare your desire to not look like a fool now vs. your desire to become better.

While one person hesitates because he feels inferior, the other is busy making mistakes and becoming superior

- [Focused Grit](#) - this basically involves sitting down, focusing and giving a problem your full attention for at least five minutes. A surprising number of problems can be solved in just five minutes. The whole process is:
  - Take a piece of paper, a pen and a timer
  - Identify a large problem, e.g. terrible job, family feud
  - Think about the problem for five whole minutes, by the clock
  - If the problem has not been resolved and you know of another 5 minute exercise that you could try, restart this process from step 1 with this exercise.
  - If the problem has not been resolved and you do not know of another 5 minute exercise you could try, spend 5 minutes brainstorming potential 5 minute exercises that could potentially solve the problem. If you come up with an exercise, restart this process from step 1.

### **Mindsets and perspectives that help you in discovering potential situations that you could end up valuing**

- Self efficacy - this is the sense that you can tackle any given problem. It is the idea that there is always a way in which to do something and that you just need to find it
- [Growth mindset](#) - this is the belief that you are always capable of being more than you currently are. A growth mindset does not necessarily involve [positive thinking](#). For example, if you repeat the affirmation 'I am a great singer'. This can actually have a negative effect. Because the affirmation relates to a fixed attribute and if your brain doesn't feel it to be true, then it is likely to activate the exact opposite concept, i.e. that you are a bad singer. Another potential negative effect of positive thinking is the "[goal-turnoff effect](#)" which means that once a pursuit attempt has been completed the goal deactivates and inhibits the mental representations used to attain the goal.
- [Mental contrasting](#) - is a technique can be used as an alternative to positive thinking and has been [shown](#) to be useful. Mental contrasting is a visualization technique where you imagine several positive aspects of completing your goal and then look at your current situation and the obstacles that are stopping you from completing the goal. To work well mental contrasting requires a growth mindset and a reasonable expectation of success. This is because if you do it



with a fixed or negative assessment of yourself then it will only deepen and reinforce this assessment.

- A feeling that you should keep [trying new things](#) - a big part of this is being willing to embrace the new and uncertain in the hope that it will lead to something useful. This [involves](#) taking on a new perspective where you don't think of planned actions in terms of their potentially aversive or pleasant outcomes, but instead as potential sources of useful information. You view them in a similar way to how you would view an experiment. If an experiment does not give you the expected result, it still gives you a range of other useful information like what doesn't work or that your beliefs may be somehow faulty. This type of thinking is not that intuitive because, by and large, the mind encourages us to behave in ways that have worked before. Though this is useful from a survival perspective, it does not often lead us to try things that are completely different from what we have done before.
- [Use curiosity](#) or closed vs open ways of thinking - for complex problems, your time is sometimes better spent, not in trying to solve the problem, but instead in searching for solutions to the problem. That is, it can be useful to jump out of your problem solving mode or your cached routine mode and to start questioning. This is especially true for problems that you have been trying to solve for a long time. It is important to note that searching for solutions isn't necessarily a wholly conscious process. It is often the case that the correct solution has already been activated by your brain, but is not being brought to your conscious attention because of other cached responses that are taking priority. If you can find some way to relax, get distracted or do something else for a while the correct solution will sometimes, seemingly out of nowhere, pop into your head.

### **Optimizing your ability to follow through with your plans**

- [Being Strategic](#) - teaches you how to make a to-do system. This includes going through tools like things like: [Getting Things Done](#) (GTD), [Remember the Milk](#), [Workflowy](#), [BeeMinder](#), [Anki](#), [Pomodoro Technique](#)
- [Next actions](#) - our motivation is normally dependent on our predictions, which can be [problematic](#). A way to avoid these problems is to just know the next physical action that would get you closer to completion. Instead of having an item 'do my taxes' you can have an item of 'find tax forms and put them on desk' which is both a concrete and small action.
- Trigger Action Plans (TAP)s - is a simple algorithm that goes, "if I do x, then I will do y." See [this post](#).
- [Structured Procrastination](#) - involves acknowledging that procrastination is somewhat inevitable which means that it can be beneficial if you spend some time trying to optimising what you will do when you are procrastinating. If you are probably going to end up procrastinating, you might as well get the most out of it. For example, you might decide to procrastinate on starting an assignment by going on a walk or cleaning the house. By using structured procrastination you can also make your plans more pliable and capable of being completed.
- [Does future me have a comparative advantage?](#) - if you are thinking of procrastinating and putting something off till later, try asking yourself if the future version of you will be better off or more easily able to handle the task. If the answer is no, then you should probably just do it now. This question is probably more helpful for small items than large ones.
- [Murphyjitsu](#) - is essentially imagining everything that could possibly go wrong and would stop you in reaching that goal. Then outplanning these potentialities so that they never occur. They is referred to as outplanning Murphy. The

“Murphyjitsu” technique can be used by asking “how surprised would I be if I failed?”, followed by “what obstacles might prevent me from finishing?” For example, “I’m going to do CrossFit three times a week and I would only be mildly surprised if I failed” should immediately lead to problems and solutions like:

- “Which days?” → “Mondays, Wednesdays, and Fridays, at 5:30 PM, after work, with an alarm set, and with my workout clothes always in my car.”
  - “What if I can’t go during gym hours?” → “If I can’t make CrossFit, then any exercise of equal or greater intensity and duration will suffice.”
  - “What if I’m sick?” → “I don’t go if I’m too sick to go to work, otherwise I go.”
  - “What if I’m, like, really really tired though?” → “I go anyway but take it easy if I need to.”
  - “What if I’m injured in a car accident?” → “Obviously don’t worry about it. Also, I need a first aid kit in my car.”
- [Habit practicing](#) - Imagine that you wanted to stop sleeping in when your alarm goes off in the morning. One way to train this habit is to practice waking up when your alarm goes off. You would do this by turning off the lights in your room, setting your alarm for five minutes, getting into bed, relaxing and then mindfully getting up straight away when the alarm goes off.
  - (I have been told that this is no longer up to date) [Urge propagation](#) - is a motivation hack which traces the causal link between a larger goal you really care about and the action you don't feel motivated to take, and explains that link to [system 1 in a language it can understand](#). It involves finding out what your System two wants you to do and how you can align your system one to want the same thing. This is normally done with [operant conditioning](#) or [reinforcement](#). What makes things like trying to lose weight or to learn the violin difficult, is that it often causes conflicts with system one driven urges, e.g. wanting a nap or craving a cookie, and because long-term goals typically require sticking it out through a series of unpleasant intermediate steps, e.g. eating less and practicing the violin, it can be easy to lose the original motivation. This loss of motivation tends to not occur so frequently for things that we feel rewarded by. This is primarily because we do these things automatically. When I want Thai food, I don't need to volitionally force myself to drive there, look at the menu, go inside, order etc. I don't have to convince myself to take those steps. They happen automatically. The point of urge propagation is to make long-term goals feel more like short-term urges which essentially means rewiring the brain to associate actions with rewards for the particular task that you want to do. You can use a similar idea to discourage bad habits. You do this by stretching out the time between an action and its reward. For example, if you want to stop reading stuff online then you can have the pages load more slowly for non work related material. Thinking about urge propagation leads us to realise that small aversive moments can have a drastic impact on our choices and motivations. The aversive moment of getting into a cold swimming pool can overwhelm the delayed rewards of doing morning laps even though you might really like or want the rewards, e.g. being fitter. In summary the process in urge propagation is:
    - Find a situation where the rewards are distanced from the actions
    - Associate the activity with a powerful feeling of reward, one with a stronger neurochemical kick than the virtuous goals, e.g. ‘being healthier, that we normally aspire to.
    - Come up with a mental image that vividly captures that feeling and that you can summon in moments of weakness. It should be a very sticky image because if it isn't then you won't experience that gut-level surge of motivation that the image is meant to cause. [For example](#), one person was



able to overcome their aversion to push-ups, which made them feel unpleasantly hot and sweaty, by tapping into their obsession with longevity. They now vividly imagine the heat from the exercise as if it was a fire that burned away cell-damaging free radicals.

## Learning optimization

- [Turbo Charged training](#) - this is best for the types of skills that you repeat over and over again to get your nervous system to change in order to get better at it. The underlying idea is the rule of intensity which states that the experience of intensity or effort that you are expending to learn something corresponds with the rate at which you are learning it. For example, 10 minutes of intense time spent trying to learn a piano piece may be just as helpful as 1 hour of less intense effort. Turbo charged training has two main principles: maximise the engagement with the learning process and minimize error. This means that if you are making lots of errors then you need to slow down or make whatever you are doing easier. If what you are doing is too easy, then you need to find some method to ramp it up and make it more difficult. It is also important that you are learning the right skills and are getting good feedback which is not noisy, but is fast enough to signal to the reward centers of the brain that you are actually making progress. This ensures that the neural patterns that were just activated get reinforced. Turbo charged training is based on taking an outside view of learning, i.e. looking at the people who learn really fast and what they are doing.

## Planning optimization

- [Surprise-o-Meter](#) - this technique involves picturing an event and observing how surprised you would be to be in that situation. Surprise is a clue that you were implicitly expecting something else to happen. The more surprised you are, the less probable your subconscious thought of that event was. Also called, [pre-Hindsight](#) which involves using emotions to evaluate how likely you are to succeed at a goal. Imagine that, six months from now, you have not achieved your goal. The level of surprise you feel at this outcome is a good predictor of whether you will actually succeed. You can then use this information as input that can be used to build plans that will actually work.
- [Your inner simulator](#) - Your "[inner simulator](#)" is CFAR's version of the distinction between [profession and anticipation](#). Basically, your "inner simulator" is the part of you that can play movies forward to determine what to anticipate: "Do I have time to turn left before that car reaches me?"; "What will she do, if I approach and say 'hi'?" You can use your inner simulator to develop high quality plans. An example of this process would be asking:
  - What you predict will happen if you take some course of action?
  - What could go wrong?
  - How can you change your plans to prevent that thing from going wrong?
  - This process is repeated from the first step with your predictions updated to take into account your preventative measures.

## Improving the accuracy of your credence levels

- [The credence calibration game](#) - when you hear someone say "I'm 90% sure...", that "90%" figure is called a credence level. Most people [tend to be over-confident by default](#), in the sense that when we say "90% sure", we tend to be right much less than 90% of the time. This game is basically getting you to practice the mental action of converting your internal sense of "sureness" into a

credence level that you can report to yourself and others. The goal is to train this over time so that your credence levels more closely reflect the actual success rates.

## **Mindfulness of how your internal state is affecting how you perceive the world**

- [Againstness](#) – is the feeling of being opposed to something. An example is when you get angry during a heated argument. This feeling typically manifests in sympathetic nervous system responses in the body. Some of these symptoms could include: hunched shoulders, rubbing the neck, positioning arms to protect the belly, raising heart rate and tensed muscles. By being mindful of your againstness you can calm yourself and engage in a [shift](#) to the parasympathetic nervous system, so from a mental point of view you might: try to appreciate it as an opportunity in practicing releasing againstness, try to increase your empathy for the other person by modelling them in enough detail that you understand why the thing that is causing your againstness seems right to them or try to emphasise your sense of tribal togetherness. This means that you should attempt to see the other person as a fellow human being and wish to help them realize their full potential. From a purely physical perspective you could:
  - open your posture, i.e. have shoulders back, spine straight & upright, head balanced on spine, belly exposed
  - breathe deeply, smoothly, gently and low in the diaphragm
  - relax the hands, arms, shoulders and eyes
  - try smiling/laughing
- [Perceptual editing](#) - is the ability:
  - to recognize when you're making a personal contribution to experience
  - to decide whether it's a contribution you actually want to make
  - to leverage the opportunity and deliberately choose what contribution you'd rather make, if any.
- [Coordinatizing emotion-space \(and feeling it\)](#) - essentially [emotional awareness](#), i.e. understanding what it feels like when you are feeling a particular emotion or set of emotions.
- [Keep your identity fluid](#) – our worldviews and beliefs affect our valuation processes. If there is something that you consider staunchly to be part of our identity then you will be more likely to believe claims related to it and to defend them. Now, this isn't necessarily a negative thing. If you want to do something, then making it part of your identity can be a great way to motivate yourself to achieve it. But, it can also lead you to become biased and to rationalize when you find things that are in opposition to your identity. A fluid identity has the discussed benefits and also avoids the negatives. A fluid identity is one that you can mindfully alter for your own benefit. Sometimes you might want to make it smaller to improve your reasoning and other times you might want to make it bigger so that it primes you to be motivated and to care.
- [Stop rationalizing: The what fooling yourself feels like game](#) – we are [prone](#) to rationalization. By practicing what it [feels](#) like when we are rationalizing we can become more likely to notice when we are rationalizing.
- [CFAR sees emotional states](#) as cues or indicators which tell you that you should double-check your reasoning or coax yourself into another emotional state. Changing your emotional state can be a potentially useful mechanism for making rationality easier. For example, getting yourself into a "playful" mode can make it easier to explore a wider set of possible actions.

## **Resolving Disagreements**

- Double Crux game - double Crux is a structured format for collaboratively finding the truth in cases where two people disagree. Instead of non-interactively offering pieces of their respective platforms, people jointly seek the actual question at the crux of the disagreement—the root uncertainty that has the potential to affect both of their beliefs. See [this post](#)

**Other stuff which has already been covered in LessWrong**

- [Bayes' theorem](#) - as described elsewhere, but more focused on the aspect of [sharing sources of disagreement and using a Bayesian framework for settling those disagreements](#).
- [Checklist of rationality habits](#).
- [Look at a problem as an outsider](#) - i.e. use reference class forecasting.

# The Brain Preservation Foundation's Small Mammalian Brain Prize won

The Brain Preservation Foundation's Small Mammalian Brain Prize has been won with fantastic preservation of a whole rabbit brain using a new fixative+slow-vitrification process.

- [BPF announcement](#) ([21CM's announcement](#))
- [evaluation](#)
- The process was published as "[Aldehyde-stabilized cryopreservation](#)", [McIntyre & Fahy 2015](#) ([mirror](#))

We describe here a new cryobiological and neurobiological technique, aldehyde-stabilized cryopreservation (ASC), which demonstrates the relevance and utility of advanced cryopreservation science for the neurobiological research community. ASC is a new brain-banking technique designed to facilitate neuroanatomic research such as connectomics research, and has the unique ability to combine stable long term ice-free sample storage with excellent anatomical resolution. To demonstrate the feasibility of ASC, we perfuse-fixed rabbit and pig brains with a glutaraldehyde-based fixative, then slowly perfused increasing concentrations of ethylene glycol over several hours in a manner similar to techniques used for whole organ cryopreservation. Once 65% w/v ethylene glycol was reached, we vitrified brains at  $-135^{\circ}\text{C}$  for indefinite long-term storage. Vitrified brains were rewarmed and the cryoprotectant removed either by perfusion or gradual diffusion from brain slices. We evaluated ASC-processed brains by electron microscopy of multiple regions across the whole brain and by Focused Ion Beam Milling and Scanning Electron Microscopy (FIB-SEM) imaging of selected brain volumes. Preservation was uniformly excellent: processes were easily traceable and synapses were crisp in both species. Aldehyde-stabilized cryopreservation has many advantages over other brain-banking techniques: chemicals are delivered via perfusion, which enables easy scaling to brains of any size; vitrification ensures that the ultrastructure of the brain will not degrade even over very long storage times; and the cryoprotectant can be removed, yielding a perfusable aldehyde-preserved brain which is suitable for a wide variety of brain assays...We have shown that both rabbit brains (10 g) and pig brains (80 g) can be preserved equally well. We do not anticipate that there will be significant barriers to preserving even larger brains such as bovine, canine, or primate brains using ASC.

(They had problems with 2 pigs and got 1 pig brain successfully cryopreserved but it wasn't part of the entry. I'm not sure why: is that because the Large Mammalian Brain Prize is not yet set up?)

- [previous discussion](#): Mikula's plastination came close but ultimately didn't seem to preserve the whole brain when applied.
- commentary: [Alcor](#), [Robin Hanson](#), [John Smart](#), [Evidence-Based Cryonics](#), [Vice](#), [Pop Sci](#)
- [donation link](#)

To summarize it, you might say that this is a hybrid of current plastination and vitrification methods, where instead of allowing slow plastination (with unknown decay & loss) or forcing fast cooling (with unknown damage and loss), a staged approach is taking: a fixative is injected into the brain first to immediately lock down all proteins and stop all decay/change, and then it is leisurely cooled down to be vitrified.

This is exciting progress because the new method may wind up preserving better than either of the parent methods, but also because it gives much greater visibility into the end-results: the aldehyde-vitrified brains can be easily scanned with electron microscopes and the results seen in high detail, showing fantastic preservation of structure, unlike regular vitrification where the scans leave opaque how good the preservation was. This opacity is one reason that as [Mike Darwin has pointed out at length on his blog](#) and [jkaufman has also noted](#) that we cannot be confident in how well ALCOR or CI's vitrification works - because if it didn't, we have little way of knowing.

EDIT: BPF's founder Ken Hayworth ([Reddit account](#)) has posted a piece, arguing that ALCOR & CI cannot be trusted to do procedures well and that future work should be done via rigorous clinical trials and only then rolled out. ["Opinion: The prize win is a vindication of the idea of cryonics, not of unaccountable cryonics service organizations"](#)

..."Should cryonics service organizations immediately start offering this new ASC procedure to their 'patients'?" My personal answer (speaking for myself, not on behalf of the BPF) has been a steadfast NO. It should be remembered that these same cryonics service organizations have been offering a different procedure for years. A procedure that was not able to demonstrate, to even my minimal expectations, preservation of the brain's neural circuitry. This result, I must say, surprised and disappointed me personally, leading me to give up my membership in one such organization and to become extremely skeptical of all since. Again, I stress, current cryonics procedures were NOT able to meet our challenge EVEN UNDER IDEAL LABORATORY CONDITIONS despite being offered to paying customers for years[1]. Should we really expect that these same organizations can now be trusted to further develop and properly implement such a new, independently-invented technique for use under non-ideal conditions?

Let's step back for a moment. A single, independently-researched, scientific publication has come out that demonstrates a method of structural brain preservation (ASC) compatible with long-term cryogenic storage in animal models (rabbit and pig) under ideal laboratory conditions (i.e. a healthy living animal immediately being perfused with fixative). Should this one paper instantly open the floodgates to human application? Under untested real-world conditions where the 'patient' is either terminally ill or already declared legally dead? Should it be performed by unlicensed persons, in unaccountable organizations, operating outside of the traditional medical establishment with its checks and balances designed to ensure high standards of quality and ethics? To me, the clear answer is NO. If this was a new drug for cancer therapy, or a new type of heart surgery, many additional steps would be expected before even clinical trials could start. Why should our expectations be any lower for this?

The fact that the ASC procedure has won the brain preservation prize should rightly be seen as a vindication of the central idea of cryonics -the brain's delicate circuitry underlying memory and personality CAN in fact be preserved indefinitely, potentially serving as a lifesaving bridge to future revival technologies. But, this

milestone should certainly not be interpreted as a vindication of the very different cryonics procedures that are practiced on human patients today. And it should not be seen as a mandate for more of the same but with an aldehyde stabilization step casually tacked on. ...

# Defiance

This is a linkpost for <https://mindingourway.com/defiance/>

The third dubious virtue is defiance. As with the other [dubious virtues](#), it can get you into trouble. [Remember the law of equal and opposite advice](#). Used correctly, it can play a key role in a healthy guilt-free motivation system.

I used to tell people that I'm roughly 90% defiance-fueled. The most common response was "ha ha I guess you can be manipulated by reverse psychology, then"; which led me to realize that I didn't yet know how to convey what I meant by "defiance fuel," so I stopped saying it. Today, we see whether I can convey what I mean by "defiance fuel" yet.

Most people I talk to about defiance think of it as a mental stance adopted against some authority figure. Perhaps they imagine a parental figure saying "finish your broccoli," and a child who hates broccoli with their jaw set and smolder in their eyes, who proceeds to eat with as much petulance as they can muster, plotting their revenge. The feeling we imagine in that child is perhaps the standard central example of "defiance."

I claim that that child does possess defiance-the-virtue, but not in their petulance, and not in their opposition to an authority figure. Defiance-the-action is in the child chewing with their mouth open in an open refusal to submit; defiance-the-virtue is in the mental actions they make *before* they start chewing with their mouth open. It's in the internal steeling they do when deciding not to be ordered around. It's in their decision to be self-reliant, it's in their refusal to take orders lying down. If these automatic and subconscious mental motions were verbalized, they might be written "I am my own person; and not beholden to your whims," or "if you push me, I push back." But they aren't verbalized, because they aren't conscious. They're reflexive.

Defiance-the-virtue is about encountering a badness that's brewing in the world, and *reflexively* doing everything you can to throw a wrench in the works, to twist things in your favor. Defiance-the-virtue is about taking nothing lying down, and refusing to let badnesses in the universe slide.

Defiance isn't about acting petulantly without hesitation: A defiant child might bide their time, knowing that if they act rashly there will be harsh consequences. Defiance is about *resisting the default state of affairs* without hesitation: A defiant child might weigh their options and bide their time, but at no point do they wonder whether they should defy. They simply dislike the situation, and so rebel against it.

Defiance-the-virtue is about having *that* reaction, to something that's wrong in the world.

---

Of course, there's an art to defying the right things. I do recommend [defying death](#); I don't recommend having the "defiance" reaction against people who tell you to do things in a stern and authoritative voice. People who order you around can either be ignored or obeyed according to the social context, but they aren't usually worth *defying*, except perhaps in situations where you legitimately need to demonstrate that you're not beholden to them, and where gentler reminders have failed.

As a rule of thumb, I suggest that it's usually healthy to have a defiance reaction towards *states of the world*, and usually unhealthy to have a defiance reaction towards *people*.

To illustrate the difference, imagine you're Neo, twenty years after the first [matrix movie](#). The sequels never happened; instead you got trapped in the matrix while one by one, all your connections to the outside world died or disappeared. One day, you lost your grasp on your ability to control the matrix, your abilities slipping through your grasp like lucidity slipping away in a dream. Now you stand atop a skyscraper, looking across the gap at its twin, unable to quite recall what it was like to fly.

You stand there frozen, desperate to recall what you once knew, finding it evasive. Behind you, someone else enters the rooftop and shouts at you over the wind.

"What the hell are you doing, you idiot?" they cry. "Get back from there! Now!"

Defiance-against-a-person would be to feel a burning need to show this person up, show them that you're not beholden to their demands, and possibly do something rash.

Defiance-against-the-world would be to hear this person cry out, and use the impetus to remember what it was you used to know. You would say, "Oh, right. I'm in the matrix." You would remember that the rules and customs of this place do not have dominion over you, no matter what illusions the people around you are taken in by. Your mind would snap back into focus. You would grab what you had forgotten how to grasp, and leap.

(And those with defiance-the-virtue deeply instilled in them don't need the impetus provided by another person to access the mental state — defiance is a property of the relationship between them and the state of the world that they can recall at will, not a property of the relationships between them and others.)

---

This is the defiance I mean to talk about. It's related to [level hopping](#) and skepticism about your limitations. It's related to the skill of measuring your progress not against others, but against what actually happens.

I've been writing a long sequence of posts on how to replace guilt-based motivation with something else. Many people have remarked to me that my writings on averting guilt seem inspired by Taoism. And: maybe. There are some parallels. But not here, not with defiance.

Defiance is not about coming to terms with the world. It's about looking looking at the world and having the same mental reflexes as the defiant child. It's about the reflexive impulse to say "screw this" and choose self-reliance over hopelessness in the face of problems that are crushingly large. It's about a deep-seated inability to go gently into that good night. It's about being able to look at the terrible social equilibria we're all trapped in and get *pissed off* — not because any individual is evil, but because almost nobody is evil and everything is broken anyway.

Above all, it's about seeing that the world is broken, and *feeling* something akin to "fuck these mortal constraints, I'm *fixing things*."

When the defiant child eats their vegetables with as much spite as is humanly possible, there was never a *thought that crossed their mind* about capitulating to their



parents. Petulance was an *automatic response*. They weren't carefully weighing a decision about whether to spite their parents — at best, they may have carefully weighed a decision about whether to get their payback now, overtly; or later, subtly. The defiance was a *reflex*; the fact that they weren't going to submit quietly to authority was never in question.

Defiance-the-virtue is about having the same reflexive response, not towards an authority figure, but towards *the state of a [broken world](#)*. It's about making the fact that you struggle to fix broken worlds *automatic and unspoken* — you might weigh your options and bide your time, but you spare no thought for *whether you will struggle*.

I don't know how to teach defiance, but it's one of the keystones of my motivation system. If you want to build yourself a motivation system akin to mine, defiance is an important component.

So this is how I suggest motivating yourself in place of guilt: Let the wrongness of the world trigger something deep inside of you, such that the question stops being whether you will capitulate or lose hope, and becomes *how you will wrest the course of the future onto a different path*. See the current state of affairs as your adversary; see the future as the prize that hangs in the balance. Shake off the illusory constraints, set your jaw, and rebel. Defy.

Allow yourself to be a little [reckless](#). Get a little [desperate](#). Let defiance of the way things are burn in you. Then *act*.

# Conclusion of the Replacing Guilt series

This is a linkpost for <https://mindingourway.com/guilt-conclusion/>

Today marks the end of my series on replacing guilt ([table of contents](#)).

I [began the series](#) by discussing the "restless guilt," that people feel when some part of them thinks they aren't doing what's important. I argued that it's possible to [care about things outside yourself](#), and things [larger than yourself](#), no matter what a nihilist tells you.

In the second arc of the series I implored readers to [drop their obligations](#) and ask themselves where they would put their efforts if there was nothing they felt they "should" be doing. If you can drop your sense of obligation and still [care hard](#) for something larger than yourself, you are well on your way to dispensing with guilt-based motivation.

In the third arc, I described techniques for building and maintaining a powerful intrinsic drive without the need to spur yourself with guilt. I point out that [working yourself ragged is not a virtue](#), and that the "work too hard then rest a long time" narrative is a [dangerous narrative](#). We can't always act as we wish we could: We're [not yet gods](#), and it's often easier to change our behavior by [exploring obstacles with experimentation and creativity](#) instead of attempting to berate and guilt ourselves into submission. I plea for [self compassion](#) and argue that [there are no "bad people"](#).

In the [fourth arc](#), I describe ways to draw on the fact that the world around you is broken as fuel for your intrinsic drive. If, when given the choice between "bad" and "worse" you can [choose "bad" without suffering](#); if you can [be content in your gambles](#) while [having no excuses](#) and [coming to terms with the fact that you may fail](#), then it becomes easy to [transmute your guilt into resolve](#) and struggle hard to make the future as bright as you can make it.

In the [fifth and final arc](#), I describe mindsets and mental stances from which guilt seems an alien concept. Primary among them are "[confidence all the way up](#)", the skill of believing in your capabilities while not being overly sure of anything; and [desperate recklessness defiance](#), the three dubious virtues of those with strong intrinsic drive.

I conclude with a few words on [how we will be measured](#): When all is said and done, Nature will not judge us by our actions; we will be measured only by what *actually happens*. Our goal, in the end, is to ensure that the timeless history of our universe is one that is filled with whatever it is we're fighting for. For me, at least, this is the underlying driver that takes the place of guilt: Once we have learned our lessons from the past, there is no reason to wrack ourselves with guilt. All we need to do, in any given moment, is look upon the actions available to us, consider, and take whichever one seems most likely to lead to a future full of light.

# Recklessness

This is a linkpost for <https://mindingourway.com/recklessness/>

The second dubious virtue is recklessness. As with [desperation](#), there are many bad ways to be reckless. There is a nihilistic recklessness, in those with a muted ability to feel and care, that is self-destructive. There is a social recklessness, when peers push each other towards doing something dangerous that none of them would do alone, in a demonstration of commitment that can become needlessly dangerous. And there is a fiery, destructive recklessness in those too quick to anger, which can lead people to actions they will regret for a lifetime. I caution against all these types of recklessness.

Nevertheless, there is a type of recklessness that is a virtue. This is *recklessness in the pursuit of an external goal*, and I have found it to be rather rare.

I get a lot of questions from people about how cautious they should be as they make changes in their lives. If they remove their guilt motivation, will they be able to do anything at all? If they really try to understand how screwed up the world is, on a gut level, will they break? If they devote their efforts to the pursuit of something larger than them, will they lose touch with their humanity, and with their ability to connect to other human beings?

And I tend to answer: You are not made of glass.

Dive in. Change things. Fix problems. If more problems crop up, fix those too.

Imagine that you look upon yourself, detect harmful guilt-based motivation, tear it out, and then notice that this leaves you with a Zen-like lack of drive, such that most of yourself is now happy to let days slip by but some small part of you is crying out that something is wrong. Recklessness-the-virtue is about being in that state and deciding to push *forward* rather than retreating; deciding to make a desperate effort to acquire a new drive, rather than panicking and retreating back towards guilt.

Recklessness is about ripping off the blinders that prevent you from [seeing the dark world](#) on a gut level, and knowing that if this happens to be debilitating then you'll find some new way to handle it, rather than being forced to retreat.

Always forward, never back. Be [unable to despair](#). Have [confidence all the way up](#). Think of all the people you know who are too stagnant, too cautious about breaking something important, to ever change at all.

You can recover from breaking a few parts of yourself, so long as you're modular rather than fragile. You can become able to roll with a few punches.

---

(This seems like a good time to insert a heavy-handed reminder about the law of [equal and opposite advice](#)! Many people would do well to gain a little recklessness, but many others need *less* recklessness and *more* caution. If you're in a particularly fragile mental state, consider disregarding this post entirely.)

---

During my undergraduate education, I was the president of an entrepreneurship club. The first most common type of person who would drop by asking for advice was that

young wannabe founder all full of naïve excitement about some half-formed notion that they're about to make the next facebook. The second most common person was that competent programmer with an idea that wasn't half-bad — maybe they had some idea for an app that would let couples communicate in a way they couldn't yet easily do, six years ago — but, being tempered and level-headed and well aware of the naïvety of the first folks, were entirely unable to *commit* to their idea.

Both sets of prospective entrepreneurs were doomed to failure. The first set, for all the obvious reasons — they'd focus too narrowly on writing code that no one would ever buy, or fail to find their first users, or fail to make a minimum testable product, or they'd dramatically misunderstand and underestimate the difficulty of the technical challenges, or whatever.

The second set would fail because they didn't really expect themselves to succeed. They could *make* themselves work on their idea, while reciting to themselves some story about being risk-loving, but they couldn't get their head *into* the idea, to the point where they were spending fourteen hours a day working feverishly while plans and paths and strategies dominated their waking thoughts.

There's a fugue state that successful entrepreneurs report entering, which the second set of people had rendered themselves unable to enter. Somehow, their realistic understating of their odds destroyed their ability to commit.

In one fashion, this makes some sense: they, knowing that great success is likely a lie, cannot fool their innermost self into believing in their own vision, which precludes them from entering the fugue state.

But in another fashion, is silly. What do the *odds* have to do with your *ability to commit*? Why is their *epistemic* state preventing them from entering the *emotional* state that would most help them succeed?

I think there are a few different skills it takes to be able to enter the fugue state even while knowing that your odds of success are low. One of them, I think, is the virtue of recklessness.

Recklessness is in the ability to say "screw the odds, I'm going to push forward on this path as hard as I can until a better path appears." If the odds are low, a better path is more likely to appear sooner rather than later — but the reckless let that be a fact about the *paths*, and they don't *further* allow low odds to prevent them from pushing forward on the best path they can currently see, as fast as possible.

If you want to become a successful entrepreneur, or if you want to succeed at other very difficult tasks, it helps to be able to take the best from both types of hopeless entrepreneurs. Become the sort of person who can enter the fugue state and give an idea your all, while *also* being able to see and avoid all the common failure modes. The fact that you are unlikely to succeed is an *epistemic* fact, you do not need to give it dominion over your *motivation*. Be a little reckless.

---

Recklessness, as a virtue, is about being able to throw caution to the wind. It's about being able to commit yourself fully to the best path before you, and then change your entire life at the drop of a pin as soon as a better path appears. It's about being free to act without worrying too much about what happens if you disrupt the status quo — too many people are already too stagnant, and we need to move faster.

So if you find yourself knowing what it is that you need to do next, but worried that doing so will break something else important...

then I say, do it.

Act.

Try not to break anything vital, but if you do, fix it and keep moving.

Always forward, never back.

Be a little reckless.

# How we will be measured

This is a linkpost for <https://mindingourway.com/how-we-will-be-measured/>

After nearly a year of writing, my "replacing guilt" sequence is coming to a close. I have just one more thing to say on the subject, by pointing out a running theme throughout the series.

When all is said and done, and Nature passes her final judgement, [you will not be measured by the number of moments in which you worked as hard as you could](#). You will not be judged by someone rooting around in your mind to see [whether you were good or bad](#). You will not be evaluated according to how [unassailable your explanations are](#), for why the things that you couldn't possibly have prevented the things that went wrong.

You will be measured only by what *actually happens*, as will we all.

That doesn't mean all of us are using the same measuring stick: Some people are working to ensure that our universe-history is one in which they in particular have a happy and fulfilled life; others are working to ensure that our universe-history is one in which their children never have to debase themselves to survive. Still others look wide, and see poverty and destitution and suffering, and work to ensure that those blemishes fade from our universe-history, in the places they can reach, near the time of their lives. Others look far forward, working to ensure that our universe-history is full of flourishing sentient civilizations and other nice things.

All it means is that the *type* of thing we're all trying to do, one way or another, is ensure that the actual history of our universe, the actual timeless structure of the place we're embedded, is as desirable as possible. That's the type of game we're playing: We manipulate universe-histories, for the sake of the future.

Some people have a listless guilt, thinking that nothing matters but feeling vaguely restless as they watch themselves spend their lives on things they think are pointless. Other people have a pointed guilt, thinking that *everything* matters, and berating themselves whenever they fall short of perfection. For me, the framing that *we act to determine the shape of our actual universe-history* is a framing that avoids both these pitfalls. Is there a way you want the completed, timeless story of our universe to go? Then act to ensure that the future is as good as you can make it. Are you wracked with guilt about your inability to act as you wish, or regret for the things you did in the past? Then act to ensure that the future is as good as you can make it. That's the sort of game we're playing: At all times, act to ensure that our future is bright.

---

I think many people get a bit mixed up about what type of game we're playing. They get stuck playing a social game, measuring their accomplishments by comparison to the accomplishments of their neighbors; or they [mistake someone else's expectations for their preferences](#) and get stuck chasing [lost purposes](#); or someone slights them and their vision narrows as their sole objective becomes *retaliation*.

I'm not saying social goals are intrinsically bad. Wealth and status are useful aids when it comes to determining the future; the accomplishments and expectations of your peers can provide useful measurements of your abilities. But there's a difference between pursuing social goals for the sake of determining the course of our universe-

history, and forgetting entirely that success is measured in terms of what actually happens throughout the course of history.

I alluded to this when I described [defiance](#) as "choosing self-reliance." At the end of the day, each and every one of us is engaged in a personal struggle to determine the future. We are not *alone*; there are many around us who can be friends and allies and support us in our struggle. But the goal, in the end, is to use what resources we have at our disposal to ensure that the universe-history is filled with light, whatever our light may be. I hope yours includes friends and family and loved ones, but *making it happen* — that is your personal task. You are encouraged to draw on the support of friends and allies where possible; and ensuring that you have close connections may be one of the properties you're putting into the timeless history of our universe: But even then, the task of *ensuring our universe-history is one in which you have close connections* is your personal task.

---

What we are doing, on this earth, is acting in such a way that our future is filled with light. From this framing, "guilt-based motivation" is a foreign concept: If you start to feel guilty, simply [look at your situation with fresh eyes](#), and then act such that the future is filled with light. Our lives are not status competitions; the world is not a proving ground. We are participating in a *gambit for the future* (or, more likely, a gambit for the shape of the multiverse), and that is all.

When there are people who oppose us out of nothing save for petty spite; when there are obstacles that stand between us and something important to us which seem all but insurmountable; when we encounter personal limitations that prevent us from acting as we wish to; it is easy to confuse retaliation, overcoming adversity, and growing stronger, with our actual goals. But crossing those hurdles is not the final objective: those hurdles are only parameters in our calculations about how to affect the future; they are nothing but the state of the game board in a game with cosmic states.

In that game, some people have stronger positions than others, and more leverage with which to determine the timeless story of our universe. Life isn't fair. But all of us, one way or another, are here to make sure that our universe history is filled with light — whatever 'light' may be to each of us.

So find allies, find friends, find everything you need to improve your ability to ensure that our universe-history tells a story you like. Move towards whatever levers on our future you can find. And then *fill it with light*.