

Beneath Psychology

Truth-Seeking as the Engine of Change

By Jimmy

beneathpsychology.com

Table of Contents

Introduction

1. [Introduction Part 1: The Challenge](#)
2. [Introduction Part 2: The Solution and What's to Come](#)

Chronic Pain Case Study

3. [Case study on chronic pain: First insights, and the remaining challenge](#)
4. [Chronic pain challenge part 2: the solution](#)

Low-Level Mechanics of Changing Minds

5. [Expectation = intention = setpoint](#)
6. [Setpoint = The experience we attend to](#)
7. [Recognizing Optimality](#)

Attention

8. [The Spectrum of Attention: From Empathy to Hypnosis](#)
9. [Navigating Attention](#)

Respect

10. [The Role of Respect: Why we inevitably appeal to authority](#)
11. [Navigating Respect: How to bid boldly, and when to humble yourself preemptively](#)

Security

12. [The necessity of security for play, and play for seeing reality](#)
13. [Navigating Security: Fighting flammability with fire \(when safe\)](#)

Clarity

14. [The Frustrations and Perils of Navigating Blind to Rocks](#)
15. [How to actually decide](#)
16. [Solving irrational fear as deciding: A worked example](#)

Tying It Together

17. [Putting It All Together: A Concrete Guide to Navigating Disagreements, and Reconnecting With Reality](#)

Limits and Failure

18. On physiological limits of sense making
19. On failure, and keeping doors open; closing thoughts

Part: Introduction

We begin with a challenge and a demonstrated solution, as we introduce the sequence. The introduction explains more about what will be conveyed and the kind of pitfalls one may run into when reading this sequence.

Introduction Part 1: The Challenge

This sequence, "Beneath Psychology: Truth-Seeking as the Engine of Change", explores how to understand and navigate psychological change from a first principles perspective. The title "Beneath Psychology" references the fact that the constraints of clear thinking lie beneath and give rise to our human-specific quirks and biases - and methods of navigating them - in the same way that in Economics, the species-agnostic constraints of supply and demand curves underlie the human-specific behavioral economics.

It turns out that setting aside worries of human-specific concerns, and addressing first the fundamental limitations, illuminates paths to resolution that otherwise remain unseen. The subset of problems we view as "psychological" are those where someone's brain isn't doing what it should. We can therefore view all of these "psychological problems" as *disagreements*. In particular, disagreements between "our" perspective and the perspective revealed by the brain (which may be our own, or someone else's) we see as "getting the wrong answer". This is a very broad sense of the term "disagreement" which covers not only disagreements over where to get dinner but also things like pain-induced suffering which can be [usefully](#) modeled as a disagreement over whether the pain is important to attend to.

Despite being long and surprisingly difficult to understand, the result of this sequence is remarkably simple. Rather than adding "techniques" and "methods" to keep track of, I will dissolve the apparent puzzles of "psychological difficulty" until it will be hard to understand how the puzzles were ever puzzling¹, and your method for "solving" these "puzzles" becomes "I just look for what is true. When I think I see it, I point". This is an *extremely* counterintuitive result, and therefore difficult to grasp let alone believe.²

It is for this reason we will begin with a demonstration in the form of a challenge:

One summer day I was sitting with my wife in the jacuzzi in our complex's pool yard, when a few kids from the local community college hop the fence for a swim. The girl of the group gets in the jacuzzi, but refuses to go swimming with her friends because she's worried that her makeup would wash off. Her friends plead with her: "Come on!", "Who cares!?", "You're fine!", but to no avail. She complains "But my makeup might wash off!", they tell her "Come on!", and it repeats.

Objectively, it doesn't matter if her makeup washes off. There's no one there that would judge her, and even if someone did judge her it'd just make them a jerk, which would be *their* problem. She doesn't see it this way, obviously.

The challenge is this: What do you say, and do, such that by the end of it she sees the truth here and is able to enjoy swimming with her friends?

How can you do it *without cheating*? No tricks³, no lying or otherwise manipulating⁴, just showing her the truth in a way that she will recognize what you say to be true *because you've justified it*. It's okay if it doesn't look like "cold facts and logic" so long as the focus is on finding *what's true*.

Keep in mind the difference between theory and reality. A lot of things that "should" work or "could" work in theory end up deflating when we soberly ask whether it *would* work, in reality. For example, you could ask her "What would happen if your makeup washed off?", but there's no guarantee she's going to want to get into an impromptu therapy session with some random guy, in front of her friends, because this random guy asked her a question that would work in the therapy room with a willing participant.

Before posting your solution, check whether you had to turn away from some aspect of reality in order to imagine it working⁵. If you were to find yourself in this situation, would you *actually* attempt your proposed solution? If she were to respond in the way she did in your head, would you feel like "Honestly, I'm kinda surprised that actually worked"? Or would you feel like "Well, I didn't see *that* coming!" if she were to not? Include your answers here alongside your solution.

The solution will come in next week's post which finishes introducing this sequence. The full explanation of where it came from and how it fits the requirements has to wait for a sequence worth of background.

For best results, put your solutions in the comments, with spoiler tags, and don't read anyone else's solution before commenting your own.

Notes

1. That doesn't mean that *all* puzzles disappear. After you get it, you won't struggle with irrational fears. You will still have to figure out which risks are worth taking, and that can be arbitrarily difficult. ↩
2. The person who served in part as a mentor for me when I was first learning this stuff, and whose insights started me down this path, now teaches therapists. He has found that even after he explains and teaches better methods resulting from the insight that much of what is sought after in therapy (e.g. relief from irrational fears) is inherently contradictory, most of his students stick to the less effective methods. They don't push back, or seem to struggle with the concept logically, but they don't connect it with "Oh this changes everything" and invest the cognitive work to restructure their understanding accordingly.

At the same time, despite kinda calling BS on him when he first told me about his revelations, I needed zero empirical evidence to be convinced. Once we role played I couldn't even *imagine* a way to hold onto the

problem. Once laid out, his solution fit so perfectly with my preexisting intuitions about how people/I would respond to the actual situation. I was so surprised that my meta level intuitions failed to predict this, that I knew his method was valid and that I needed to get to work recalibrating my meta-intuitions. It was still possible for it to empirically fail, but given what I already knew, *that* would be the surprise requiring evidence to believe.

This sequence follows the same pattern, and will be counter to meta intuitions while fitting object level intuitions; by design. ↩

3. "No tricks" meaning "Can't offer to pay her to get in the pool", "Can't flick salsa on her face so she has to wash it off anyway", or anything else that dodges the point of the exercise. ↩
4. "No manipulating" in that your focus must be genuinely on collaborative truth seeking. E.g. No "I hypnotized her, lol", and no "I distracted her from that idea and told some jokes so she was in a better mood and less focused on her makeup". If her explanation of why she went swimming is anything other than "It's okay if my makeup washes off", or if your honest explanation is any different, then it doesn't count. ↩
5. This distinction can be easy to miss, so I'll give an example to illustrate the difference.

The other day my 6 year old daughter was playing The Secret of Monkey Island and needed to find some meat in the game. The hint I gave her was to think about where *she* would go to find meat - because that's where it was in the game. "Hunting!". Something easier. "The store!".

These are both understandable responses, but they're both wrong. I know that's not where she would go to find meat because when I told her "Scavenger hunt time! *Go find meat and bring it to me*" she didn't start gearing up to hunt the squirrel in our backyard, and she didn't ask me how to get to the store. She immediately decided to go the kitchen and check the fridge.

This failure to distinguish between where "a person" can go to obtain meat and where *you* can go to find meat, is the exact failure I'm attempting to warn against here. The point of the challenge isn't to infer my hidden assumptions and come up with a password that works within a fictional simplification I have in my head. Nor is it to find an authoritative simplification from Psychology so that you can lean on that authority to justify your chosen solution for what "a person" can do - regardless of whether it fits reality or not. The point is not to find something that sounds good from the chair you currently sit in.

It's to find something that sounds good *from the jacuzzi*, when your eyes are open, and in contact with hers.

Because if you ever find yourself in a situation like this, that's where you'll be sitting when you're saying it.

↩

Introduction Part 2: The Solution and What's to Come

In the [opening post](#) we introduced a challenge in which you're sitting in a jacuzzi and you have the opportunity to convince a stranger that she doesn't have to worry about her makeup washing off, so she can go swimming with her friends. It was framed as a *rationality* problem, where the goal is to convince her by justifying that the idea that she doesn't have to worry is *true*.

The solution here, as with the solution to any other "psychological problem" that is really a disagreement, relies on not only having the correct answer (e.g. "It doesn't matter what other people think"), but also the justification that allows you to *know* it to be true -- and that the person will be better rather than worse off for daring to engage with your perspective.

Knowing something to be true, in the face of disagreement, requires understanding your counterparty's perspective well enough to know that your own holds up in face of theirs. Empathy is *critical*, for the same reason that maps are critical when traveling; if you can't see the territory, you're SOL. I had to correctly infer what her actual concerns were, and show her that the union of my perspective and her perspective was indeed my perspective. If I can't do that, my options are to a) say stuff that predictably won't get her in the pool ("C'mon!"/"You're fine!"), and show myself to be unable to face reality, or b) humble myself, recognize that I actually don't know better than she does, and don't try to change her mind because I don't actually know the right answer.

If I *do* know what it's like to be an insecure girl, well enough to recognize what her reasons are for her beliefs *and* where her perspective falls short, then I can demonstrate that. I can show her that I'm not interested in giving advice that fails to take into account her perspective, that I *do* understand her perspective, and that I have a perspective worth considering. In other words I can earn her security, her respect, and her attention.

Here's how I did it, in this case:

The first thing I did, was to interject with "Yeah, don't do it. Your makeup might wash off, and then people would see what you really look like."

Take a moment to notice how you would anticipate her responding to this. Measure your own anticipation of what kind of response is likely against what actually happened, and notice whether her response is actually surprising, or expected.

She was taken aback, and not sure how to take my comment. She challenged me with "Are you calling me ugly!?"

Can you predict how I will respond to this? Confirm intent to offend by saying "Yes"? Back pedal with "No"? Something else?

I responded "No, of course not -- I have *no idea* what you look like under that mask. I just think if you're worried, it's probably for a good reason... right?"

Do you think she said "Yes, of course I have a good reason. I am ugly. That's why I took offense instead of saying "I feel seen""?

Do you think she said "No, of course I don't have a good reason for my feelings. Where would you get such a silly idea? You can tell *I* don't take my feelings seriously, right?"

Neither fit, right? Can you sense how she *would* respond to such a question? Does it seem like maybe even *she* doesn't know?

What do you think happened next?

At this point, she was in full "WTF!?" mode. But she was smiling ear to ear. I had just *shown* her that she didn't need to be insecure, and that it was okay to play, and to lighten up a little. She asked who I was, I introduced myself, and after taking a moment to process the situation, and laughing at it all, she wanted my actual input. "I think it *doesn't matter* what you look like. If someone wants to be a judgy jerk about it, that's their problem".

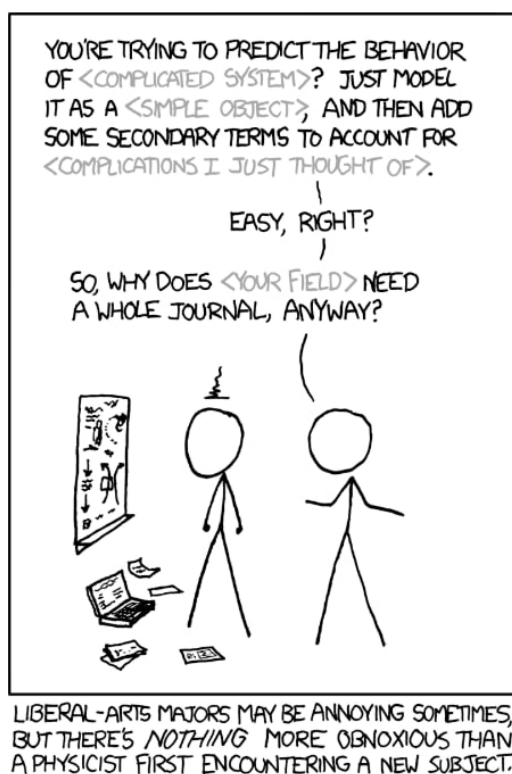
You might struggle to articulate *why* that worked, how I *knew* those responses would go well, and how I came up with them in the first place. But you can probably sense that *this works*. You can probably tell that her responses fit what I said to her, and you can probably intuit that she got in the pool after that. Furthermore, you can probably tell that the reason it works has something to do with the fact that when her friends say "It doesn't matter what people will think!", it wasn't *credible*. Yet after some friendly trolling *it somehow was*. For whatever reason, the latter is a lot harder to fake than the former¹.

And that matters.

This demonstration makes for a quick and striking example of the full complexity of navigating disagreements, and has most of the elements that will be described in this sequence. I wanted to start with a demonstration that you can evaluate for yourself, because without evidence that this isn't just theoretical, a lot of what I say might have provoked *too much* disbelief². But explaining what is going on here and how I was able to find that solution in real time will require a lot of background. Before we dive deeper into these principles, let me share how I came to this perspective.

My degrees are in physics and electrical engineering, not psychology. My day job (and hobby since high school) is as an R&D engineer where every project I work on involves a first

principles big picture approach, which shows everyone to be missing the important piece of the puzzle. Often this results in stupid obvious ideas like "There is an optimal number of switches. What if we used that amount", which yields an order of magnitude lower loss than people expect for "high efficiency aircraft inverters"³⁴.



<https://xkcd.com/793/>

(In all seriousness, I did put a lot of work into this over a decade or so, so the flippancy doesn't really fit)

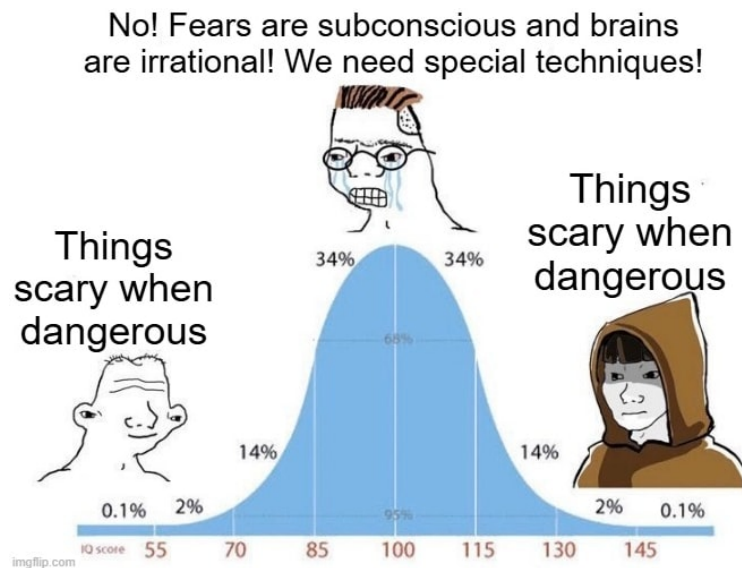
My interest in the "psychological" started back in 2010 when I realized that my belief of "hypnosis is obviously BS" was actually unfounded, and that there was no good reason for a strong prior that our brains aren't incredibly "hackable". I poured over all of the science⁵, immersed myself in forums of likeminded explorers, absorbed everything I could from what was currently known... and then with the help of one of my physics friends, made a chatbot which hypnotized thousands of people⁶ to allow me to perform double blind tests of specific fine-grained techniques.

Perhaps surprisingly, or perhaps not, what ended up being far more useful than the hypnobot was doing it myself, with deliberate attention on not only "what worked" on some sort of zoomed out perspective, but what worked *at each step* of each interaction, and *what it felt like from the*

inside, from the perspective of the person on the other side. I applied methods of thinking I had learned in physics, engineering, and the sequences to the "soft science" of psychology with an expectation of concrete results and short feedback loops. Over time, though accumulation of perspective shifting experiences, the underlying structure started to become clear.

As it did, I became less reliant on the 'falsework' of hypnotic techniques. It became less and less necessary to micromanage the lower levels of abstraction, until I no longer bothered with the trappings like "spirals and hypnotist's cloak" (alright, actually like "inductions" and "convincers"), and just focused on the structure of belief itself. Rather than coming up with anything new, I ended up returning to methods we all use for most things most of the time. Normally when we're afraid of things its because we think they're dangerous, and seeing that they aren't alleviates the fear. Sometimes, when our thinking becomes clouded, we think the normal rules have failed and invent epicycles to explain the failure instead of recognizing the failure to be ours ("this fear is 'irrational', because it didn't listen to me and there's no way *that* is rational!"/"It's 'subconscious' or something, because I can't be wrong about what *I* believe!").

I don't offer a "new method", just a more explicit explanation of what we're already doing when things work, why "solving epicycles" is necessarily worse, and how we can get all the results we can hope for by being *less wrong*.



What kind of thing is this framework

As alluded to earlier, the framework I present in this sequence is a different kind of thing than the "methods" or "techniques" you might think to compare it to. In contrast to things like CBT/ double crux/etc, the goal is to *avoid* making claims about what is causing these "disagreements"

or otherwise placing requirements about what *must* be causing them in order for a suggested response to be effective. Instead, the goal is to lay out the space of possibilities for what *can* cause disagreements, the potential paths through which one can move, and transfer much of the empirical prediction from the framework itself to the individual using the framework to navigate a particular disagreement. Rather than "My new technique didn't work, now what?", it's "I was wrong about what this situation is. In what way was I wrong?" What I aim to convey is more akin to a coordinate system and equations of motion than a solution to any particular set of initial conditions. The scope of applicability is far larger, but computing the solution is on you.⁷

The really peculiar thing is just *how much* this simplified things, and how much more effective it is than even hypnosis *at its best*.

For example, when I was first learning hypnosis my kid cousin got hit in the face with a basketball and was crying. I decided to try my hand at conversational hypnosis, to see if I could make this stuff actually work.

You know what it's like when you go to the dentist to get a filling... and they give you shots in your mouth so that it gets all numb and you can't feel anything? You're starting to feel that a little now, aren't you? That's right, now just let that feeling get stronger, and let that numbness spread up your face to your nose...

In thirty seconds I was able to numb away the pain, leaving him happy and eager to get back out to play. I thought that was *super cool, super exciting*, and if you would have told me that 1) I had no idea what I was doing, and that 2) that intervention was an order of magnitude *too slow*, I wouldn't have been able to understand how that could possibly be true.

Yet a few years later, after I had learned more, when that same kid had picked up the wrong end of a fire poker and was sitting there fighting back tears, I realized that I actually had no idea how to deal with situations like this. I viewed my previous "solution" as *not right*, and I had no idea what to do. Not knowing what to do, I turned to face that uncertainty.

And resolved the suffering in a single word.

It took *one single word* to transform his experience from "trying to hold back tears because it hurts" to smiling because it "Doesn't hurt, just tickles".

What was the word, how was it delivered, and why did it work? What has to be done such that this kind of thing can work more often?

(just space in case you want to try to solve it before seeing the solution)

The word was "Hurts?", as in "[What's wrong? Is the problem that it] Hurts?".

And the reason that it worked was that there quite literally was no problem, except for the fact that no one had noticed that there was no problem. His answer was "Yeah!" as if someone finally understood, only to immediately cut himself off with "Actually, no. When my brain feels pain it just interprets it as tickles". We'll go over this a bit more later because this definitely isn't a common reaction, but think about what this looks like from the perspective of the woman who showed up with me. She saw the kid in pain, she saw me squat down to his level, look him in the eye, and ask if it hurts -- and the kid said "Actually, no".

It's not even that what I did was "obvious in hindsight", in the sense of "I should have thought of that!" It's that it's so natural that anyone *could have* done it. It didn't require any weird "imagination games", just checking in with the kid. And so naturally, the woman who showed up with me didn't notice that I did anything unusual. Until I pointed out that even though anyone *could have* done it, no one else *did*. There were about a dozen people there trying to help the kid feel better, and despite⁸ their best efforts the kid was suffering until we showed up and I asked.

The reason *I* thought to ask, is that studying and practicing these things had taught me that pain literally can't be a problem⁹. Pain feels like it's more than "just information" because it's information *that matters* -- at least, so far as we can tell -- but it's still just information. And information itself can't be a problem, yet he seemed to be responding to it as if it were. Which is weird. So it occurred to me to ask.

And it didn't occur to anyone else there, because no one else saw anything weird about him suffering given the obvious source of pain.

What this sequence will give you

To the extent I succeed in conveying these ideas, you'll experience a shift in how you categorize challenges. You'll no longer struggle with "irrationality", and your obviously irrational beliefs and behaviors will melt away. "I have an irrational fear" becomes "Is it actually dangerous?" -- and when the answer is "No" you won't be afraid. You'll be left with the fears that *aren't* obviously irrational, and it will be a disappointingly large fraction of the ones you were sure were irrational. This reframing is generally humbling like that; it becomes harder to believe we're definitely right once we see the paths from "definitely right" to successful interventions, and can't always take them. But hey, it gives us paths to successful interventions to the extent that we're actually justified, so it's a good trade. It's really nice knowing how to reason with the "unreasonable", should it become worth the effort. And how to invest the effort pre-emptively, so repeated interactions just get easy.

You'll also learn how to try *on purpose* -- something that few people know how to do, and few people realize that they don't know how to do. This, combined with a widened view of behaviors potentially under your control, can lead to ability to exert control over "unconscious physiological stuff" like vasoconstriction, edema, [immune system response](#)¹⁰. It will also lead to the ability to deliberately recontextualize pain to where it doesn't cause suffering, and suffering to where it can be something you *choose*.

"To the extent to which I succeed" is an important caveat here. This stuff is really difficult to teach for like.... a bunch of reasons¹¹¹²¹³¹⁴¹⁵¹⁶. I'll put most of them in footnotes to make a linear read through more bearable, but they're all important points.

The one that I won't relegate to a footnote is that the things I'm describing tend to feel impossible until they feel obvious, with little in between; mundane solutions to magical problems. This isn't just a prediction, this is repeated observation¹⁷¹⁸¹⁹. The "one word pain cure" is set up to highlight this so that you notice the pattern and don't fall into this trap. There will generally not be moments of "Oh that's clever. I never would have thought of that" -- just "Oh that's impossible" to "That's not even impressive"; the solutions don't get better, the problems just seem easier.

A realistic expectation, and my hope in writing this, is that this sequence will convey enough understanding to make significant improvement consciously navigating some clear inefficiencies, and serve as a pointer that will help those motivated to discover a fuller understanding for themselves. In a previous attempt to convey these ideas, I have managed some degree of direct translation from understanding to results *without* conscious effort, similar to the effects people report with "The Easy Way to Stop Smoking" by Alan Carr and "Healing Back Pain" by Dr. John E. Sarno. Specifically, my friend noticed herself behaving differently after reading the "respect" section in the aforementioned previous attempt to convey these ideas, and believes that change directly contributed to her getting tenure²⁰. However, if it ends up coming across that well here it will have exceeded my expectations.

Brief overview of the framework

As explained in the first post, psychological issues can be modeled as disagreements. These "disagreements" can then be navigated to resolution by negotiations on three layers, which will be discussed in depth later. Briefly,

- Attention is the driver of all change. Results follow iff attention is directed in the necessary place, screening off everything else.
- Respect is necessary in order to bid for attention. Attention must be negotiated, and respect determines what you can ask without being laughed off.

- Security is necessary in order to bid for respect. Respect must also be negotiated, and security determines how boldly you can bid for respect without triggering fight or flight

Clarity forms the foundation that makes security, respect, and attention sustainable. You can fake it, but without a clear grounding it's only a matter of time before your credit runs out and you're left to pay your debts.

Outline

To develop this framework from the ground up, I plan for this sequence to include about nineteen posts (totaling approximately 70k words), published nominally every two weeks to give time for discussion and questions between posts.

First, we will start with another challenge, and a demonstration in all the messy details it takes to deal with something like debilitating chronic pain. (Two posts)

Then, we'll lay out the foundations (three posts):

- Expectation is equivalent to intent, and functions as the set point to which we control.
- Our set point is the experiences we attend to.
- Optimality looks like focusing on uncertainty about what we want to achieve, and then on the answers we want to realize.

Next, we'll explore the core framework layer by layer (six posts):

- Attention on the spectrum from empathy to hypnosis, and how to choose what's appropriate.
- Respect as the credit that allows us to communicate across non-trivial inferential distance.
- Security as a prerequisite to play, and play as necessary for growth.

Then we will explore some ways in which lack of clarity systematically misleads us into poor effectiveness -- or worse, misaligned effectiveness (three posts).

Then we will tie it all together, and show the fundamental questions holding the uncertainty which lets us negotiate on each layer, and find the appropriate paths towards resolution (1 post)

Finally, we will close with thoughts on physiological and epistemic limitations (two posts), followed by an open thread with some open ended exercises for the reader, and a space for people to ask about how to apply this thinking to the scenarios they're interested in.

Notes

1. Not impossible, as we will see later. But certainly much harder and more costly. ↩
2. Disbelief becomes "too much" when it goes beyond mere lack-of-belief into dismissing the evidence and rationalizing not having to track it. Initial disbelief is expected, and healthy skepticism encouraged. So long as you notice what the actual claims are and when little bits of evidence are beginning to add up, that's enough. ↩
3. I think they were expecting 97-99%, if I remember correctly. We were getting 99.8-99.9% depending on conditions. The "optimum amount" changes with technology/application/etc, and these problems don't stay unsolved forever. ↩
4. Or in the case of my high school project "our power factor is on the order of 10%. Why not double up on the resonance trick and bring it to one?", leading to the drive topology used in those tesla coils that make music with large lightning bolts (like the ones you might have seen [depicted in The Sorcerer's Apprentice](#) and [done for realsies with David Blaine's performance](#)). ↩
5. [The Oxford Handbook of Hypnosis](#) is a good place to start. ↩
6. The final number was around 10k, with about three times that many total conversations. ↩
7. Expanding on this a bit, by removing all predictions that can be false and making sure we can "explain everything", we work simultaneously towards "unfalsifiable" and "tautologically true". The risk with such frameworks isn't that they might be false -- because again, they don't say anything that might be wrong -- but that they might be *useless* because they don't make predictions at all and therefore offer no object level guidance.

The usefulness comes to the extent it helps us make object level decisions that aren't predictably wrong *according to our own beliefs*, and gives us a way to notice when we have beliefs pointing us in opposing directions. This turns out to be extremely common. ↩
8. And *because* of.

They were saying things like "Oh man, that must hurt!", "Let me get you some ice!", and "When I burned myself like that, I was suffering all through the night and couldn't sleep."

They were essentially unwittingly gaslighting the poor kid into suffering. ↩
9. Specifically, it was following [this 'acknowledgement' protocol](#) when I injured my foot that gave me this realization. The technique worked very well and *completely* resolved my suffering. It was really bizarre trying to figure out exactly when the pain stopped being a problem, because you'd think you'd notice when something like that happens. Turns out, that's a trick question.

More on this experience in the chronic pain post. ↩
10. This one in particular turned out surprisingly simple for me, and didn't require any special insight. I just visualized my body attacking the wart with tiny little soldiers, and within two to three days I was in disbelief because it sure looked smaller. After a few weeks, it was gone. ↩
11. There is much to say, and I can't say it all here. This sequence will err on the side of including concepts with unclear connections and justifications, as an alternative to omission when full elaboration would be prohibitive. When I wrote my "maximally terse" version for my own understanding it ended up at 34k words and the feedback I got from the one other person I expected to understand (who shares a lot of the background and helped shape my thinking on these things) was that it's way too terse and that he was going to need a second read to make sense of what I was saying. ↩

12. I think one of the big difficulties with understanding the ideas I convey in this framework is that it's legitimately subtle, and therefore easy to miss. When you're doing things 99% right, and only 1% wrong, it can *easily* feel like you're doing "good enough". It can be very tempting to cut yourself some slack, let yourself off the hook, and round up to one. The problem is that when you build on these results and iterate hundreds of times, the difference between 1^{200} and 0.99^{200} is the difference between night and day. The difference between "no obvious insecurity" and "no insecurity" can be really shocking, and hard to notice until you actually have an example of the latter to compare to -- and notice the reason for the difference. ↩
13. Big shifts in framework are hard to grapple with. As mentioned in the opening post, therapists in training tend to blankface when presented with better methods that involve seeing desire for (some) therapy as inherently paradoxical. I got largely the same response to my Tesla coil idea, and it only became popular after my Tesla coiler friend stumbled upon the same idea and made it easy to replicate by providing schematics of his version to copy. ↩
14. The meaning of a belief is in [what experiences it leads you to anticipate](#), and it often takes quite a few iterations of trying to apply these ideas before it's clear what to anticipate as a result. ↩
15. Insecurity is a bitch. We often get sensitive about our failures and in particular our "irrationalities", which we are full of. The idea that the solution was stupid simple all along can make us feel dumb, and so can the idea that our failures were actually our fault all along. That can cause us to flinch from finding out if we're not careful (i.e. security failure); so if this sequence is useful to you, it might also evoke negative emotions at some point. Rest assured, if this makes you dumb it makes me dumb too -- I fell for all the same mistakes everyone else does, and it took me a long time and a lot of work to figure all this out. I don't expect this to hit everyone, but some. ↩
16. The second and last reason compound. It doesn't take much to derail things, when you don't recognize it to be important. ↩
17. One of my proofreaders had this to say: Well I do like the paragraphs where you explicitly call out that it feels impossible then easy and no in between. Because that's all very very true. I noticed that in the book iteration a couple times where I thought why was he belaboring this point that's fucking obvious. Then remembered the many pages of Google doc of me being all "wtf that doesn't make sense". I don't remember specifically what that was on but it wasn't a singular occurrence. ↩
18. I've had someone get obviously upset claiming that I'm making ridiculous and untrue statements and then pivot without shred of irony into saying that my claims weren't novel or impressive. This person isn't exactly a shining example of a skilled rationalist, and people here do tend to be more epistemically cautious than that, but he wasn't dumb either and the content is definitely confusing in this way. ↩
19. You can play with this yourself, after I post the chronic pain transcript. If you ask an LLM how surprising the results are, and to what degree the pain was "physical" vs "emotional", you'll get different answers depending on whether you've shared the solution or just the setup. ↩
20. This is a good example of the kind of change that can be easy to miss. She noticed "from the outside" that her behavior was changing and she was getting different results, but from the inside, the way she generated behavior still felt like "just responding with what feels appropriate". The change was just a better grounding for her sense of how much tension is okay and appropriate. Because this isn't something that's normally consciously thought through, she never would have noticed from the inside alone. ↩

Part: Chronic Pain Case Study

The next two posts pose another challenge, and show the framework in action in a messy and high stakes real world application.

Case study on chronic pain: First insights, and the remaining challenge

In [the last post](#) I took the seemingly-naïve stance that "pain is just information" and "can't actually be a problem", implying that painful situations can be dealt with without suffering by just looking through the pain towards the reality at which it points. I did give an example of how this allowed me to quickly resolve an episode of pain-induced-suffering when others couldn't, but that one is admittedly a bit of an outlier and most cases aren't so simple. Today, we're going to start looking at a case that isn't so simple.

I was talking to someone through private messages on a forum for a shared hobby, and his chronic pain issue came up. Rather than acute pain from a burned hand, this was incessant pain from a medical procedure gone wrong a year prior. This procedure resulted in nerve damage, and therefore pain that feels "like every kind of pain all at the same time". The pain was so intense in the first day that the nurse was crying from seeing how much he hurt. They ended up prescribing him so much pain killers that his teeth literally started falling apart. Even still, that amount of pain drugs was only enough "for a little bit until [his] body got used to it".

The question, is why he didn't just respond to the pain by saying "Oh wow, something is wrong. Okay. *What's wrong?*" -- and then, once he figured out the answer to that, pivoting to "Ooh, neat sensations that don't mean much of anything. Anyway..."

The obvious but mostly worthless answer is "Because it hurts!". It's easy to empathize because we've all been there, but until we learn to square that experience with the reality that we don't have to suffer just because of pain, we're not going to be able to help people reconnect to reality -- and might be those people unwittingly gas-lighting others into suffering unnecessarily.

So what *exactly* is stopping him (and ourselves, when we feel similarly) from simply facing reality? Where does this desire for pain killers come from, given that they aren't exactly healing his body? Why is he suffering, rather than simply mourning loss?

What do we have to say, and how do we have to say it, such that we can navigate from here to where he *can* and *does* simply face reality? Until we can use our understandings to navigate successfully, they're not worth much. So how do we do that?

Like the last challenge, we know that it's a solvable problem because it was in fact solved. At least, partly. He described the change as "a huge improvement" and "Light-years ahead of where I was". His pain killer use dropped overnight from "norco+crazy strong stuff 3x/day" to "norco and the crazy strong stuff at most once a day". I didn't get a complete resolution here, but that

has a lot to do with the fact that he lost motivation to work further on reducing his suffering, and instead started doing things that made the pain worse -- but the reality better, in expectation.

First, I'll share the beginnings of the transcript, then give some more hints about what to look for and where to look. The neat thing about this one is that because the entire intervention is in text, I can share essentially all of it and there's no hidden body language working behind the scenes.

The following transcript has been lightly edited for obvious typos and had some parts snipped/redacted for anonymity/privacy, as well as pulling out the interspersed unrelated conversation about our shared hobby. Nothing has been reworded or summarized, and what is left represents and essentially complete and raw depiction of the whole thing.

OCT 11

other-guy: I used to hike in the mountains and camp it bushcraft all the time. Bring a backpack and sleep under a tarp, maybe build a raised bed. Until I got sick.

Jimmy: What kind of sick?

other-guy: My doctor was giving me an infusion to treat my rheumatoid arthritis, and I had a terrible reaction to it. Put my whole body in the worst pain ever and affected my muscles. I had a hard time moving my arms, and my legs became really weak, so it's really hard for me to walk now. I can use my arms better, but sometimes it's like my mind won't connect with them. Lost about 20lbs of muscle in almost two weeks. Couldn't work because of it, so that's why I'm broke, and I just keep going to physical therapy to try and get better. It's been a long battle.

Jimmy: Jesus.

other-guy: Sorry for the long story haha. They say it will take up to 3-3.5 years to completely heal. If I heal. So far, I'm at just over a year.

Jimmy: The story is short, but holy shit, that sucks. It's like getting hit by a bus, only without ever doing something stupid like standing in the street without looking. At least there's a likely healing in 3-3.5 years.

other-guy: That's honestly the worst part. If I was speeding or something and crashed, like you said, doing something stupid, then I could accept it. But I just did what the doc told me. Some parts of it I will probably have forever, like the dystrophy. It triggered myotonic dystrophy that basically attacks muscles, so I will always have to watch my heart to make sure it's not attacking it. Same with my lungs and eyes. Most of the pain should go away though. They're not sure about the mind-to-muscle connection though, which causes the trouble moving sometimes. The meds they gave me to treat everything

affected my teeth, so I had to get some of them pulled and replaced. It just kind of affected everything.

Jimmy: Not that this one would have been avoidable, but IME "doing what the doc said" isn't as safe as you'd think it should be. I'd have a couple pretty significant medical issues gone much worse if I had done what doctors said all the time. For future reference, "Uptodate" is where doctors are supposed to look to stay on top of the research, and there are Iranian sites that will allow you access for free. For example, here's one that might be relevant to you: <https://medilib.ir/uptodate/show/7480>

Jimmy: ChatGPT is also pretty useful for figuring out what to even search for and summarizing things. Here's one it found based on your description of the problem: <https://medilib.ir/uptodate/show/7501>

OCT 12

other-guy: Wow, that's amazing! Thank you! I had no idea you could do that. I'm gonna be digging into all that for sure. Thank you. Docs wanted to give me a different infusion to try and help, and I'm like fk off, no way. And they want to do spinal surgery to attach some wires that shock and disrupt the pain signal to my brain so I wouldn't have the pain. I didn't trust that either. Even the pain pills they gave me took the pain away for a little bit until my body got used to it, and they won't prescribe anything else. But that's what caused the teeth problems—changed the pH in my saliva or something that made some of them chip and expose the pulp (pretty much the nerve), and I had to get new teeth. Even though I brush 3x a day and only had 2 cavities in my life, the stuff they gave me ruined them. I'm really tired of doctors at this point.

other-guy: It's crazy how relevant those two are to my situation. Thank you again. 🙏
This is great!

Jimmy: Hopefully they help. That resource was a godsend when my wife was dealing with her pregnancy complications, so now I tell everyone I can. It'll shut you out after four articles, but it just takes clearing cookies, then you can read more. I just open a new incognito window every time to make it easy.

Jimmy: What's the pain issue, exactly? What happens if you don't take the pain meds?

OCT 13

other-guy: Good to know. Thank you. Idk how to clear cookies, but incognito I can do haha. If I don't take them, then pain from the Parsonage-Turner syndrome it caused gets a lot worse. It's basically a pain in my chest, almost armpit area, and down my arm into my hand that feels like road rash or like I burnt my whole arm. Pain from drug-induced neuropathy gets worse—it's like pins and needles everywhere but way worse than when

your foot falls asleep—and mostly a deep muscle pain becomes terrible in my legs and arms. It's like when you're working out and trying to get one more rep in, but the muscle hurts so bad like it's gonna tear or pop.

other-guy: All of this pain is constant. I feel it all the time, but it's kind of dull. It's bearable at least. If I miss a dose of pain meds, then it gets terrible, and I almost panic because it's like IDK what to do to get rid of it. The day this happened, I went into the ER, and they kept me overnight. They gave me five doses of Dilaudid (I think that's how you spell it), and one of the nurses was crying because she didn't know how to help. Nothing was working. I couldn't stay still from the pain; I couldn't focus; I kept grunting and screaming. That first day was the worst pain I've ever felt. I would rather deal with a perforated ulcer again and have surgery for it again instead of being in that pain again.

Jimmy: Interesting. It sounds like that infusion really fucked with your nerves.

Jimmy: Between the "mind-muscle disconnect" and the overwhelming pins-and-needles thing.

other-guy: In a couple weeks, I'm having a test done to test for permanent nerve damage. I also get these jolts or twitches. Like sitting down, I'll all of a sudden kick my leg up, or I've actually smacked my face when my arm twitched while looking at my phone haha. I cut myself all the time cutting food from random twitches. It's the worst at night, trying to fall asleep, and I'll get full-body ones that hurt. It's like when someone gets shocked in the hospital on TV—my whole body jolts up.

other-guy: Man, I'm so sorry to share so much with you. Besides doctors, I haven't told anyone this much. Just tell my friends I'm having troubles, and family knows a little bit more but not much.

Jimmy: How do they test for "permanent"?

other-guy: They will do some sort of EMG. They will poke me with a bunch of needles that test for electricity in the muscles, I believe. Something like that. I had one done about a year ago, and this one is to see if anything is better, worse, or the same. It's not the most enjoyable test in the world haha.

Jimmy: I've had that test, I think. Might be somewhat different, but poked with needles listening for electrical activity in muscles.

Jimmy: That was related to the first time I'm really glad I didn't do what the doctor said I should do. I had a bulging disc pushing on my spinal cord, and the doctor wanted to do a total disc replacement (at like... 20).

Jimmy: The second opinion I got was literally "that's crazy!" So I went with the second surgeon and walked out of that surgery the same morning.

other-guy: Wow, he probably suggested that just for more money. It's hard to trust anything they say these days. The Dr. that told me to get the infusions literally dropped me as a patient after it happened. Prescribed something, told her it's not working, and she said, "Well, it should be." I told her it's not, I need something else, and she dropped me, said I was too difficult, and canceled my appointments with her.

other-guy: Took me about four months to get into another Dr. Took another 4 or 5 to figure out what was going on, and new things keep happening. Didn't get into the pain clinic until maybe six months ago. That helped a lot at first, but now I'm used to the pain meds, so they're not as effective. He won't prescribe anything else, and IDK how to ask him for something new in fear of sounding like a druggie. Sorry, back on subject. What caused the bulging disc?

Jimmy: That's a good question, and I don't have a short/good answer. It runs in the family, though. My dad and brother had the same surgery. Grandpa had back issues too, but that surgery just fixed it.

Jimmy: I was back wrestling by the end of the quarter, and I've had essentially no back issues since. Dealing with doctors has been very frustrating for me in general though. I'm not sure if it was money for that guy (I'm sure it didn't help) or just standard incompetence.

Jimmy: I've had to humiliate a couple doctors by rubbing their noses in the data to get them to do what was needed even without going against their monetary interests.

other-guy: One dentist I went to said we have to pull all the bad teeth and do implants. Whole thing would be 16k about. Went to another dentist, and he said save the teeth, just need to do crowns. Total is about 6k. That's including the deep laser cleaning.

other-guy: I think doctors are like the Internet. Search your symptoms, and you find 15 different answers. Go to 15 different doctors and probably get 10-15 different answers. I'm glad your back is better. That's not something to take lightly. One thing I can't stand is back pain and back problems. Every single movement triggers the pain, and a lot of times resting only does so much, barely takes away the pain.

Jimmy: Heh. I haven't had so much issue with dentists, but mostly because I don't go much I guess.

Jimmy: I didn't go for 15 years, and the dentist was freaking out, saying my mouth was going to be full of cavities. He only found 2, and then I didn't go back to get them filled

for like a year, and then a different dentist at that place said, "Eh, I wouldn't even call those cavities," and didn't recommend fillings.

Jimmy: And then again with another not-quite-filling-like thing that they didn't see by the time I went back in. So my approach to dentists has kinda become "just wait until they forget about the problem, and see if they see it again next time."

Jimmy: Is your back pain related to this infusion thing, or just a separate thing you've had to deal with before?

OCT 14

other-guy: Dentists are new to me. Only had a couple cavities as a kid, not since high school until all this crap. The back pain is unrelated. I don't get it often either. I used to lift a lot and was lifting really heavy. Something happened to the bench, and it slipped out of the angle it was set at, causing me to drop the weight but tore both rotator cuffs and something to my back.

other-guy: It barely bothers me anymore, but when it does, I absolutely hate it. The rotator cuffs bother me more than my back does.

Jimmy: Geez, that sucks.

Okay, so how do we go about this?

The insight that suffering is distinct from pain is important, but what happens if you just tell him "Suffering is distinct from pain"? What if you tell him "You need to accept the pain, then you will no longer suffer!"?

I think it's pretty clear to everyone that "accept the pain" is easier said than done, and that "intellectually knowing"¹ that suffering is eliminated by acceptance is not in general sufficient.

It's insufficient because it presupposes that he *should* stop suffering. What if we don't presuppose that he should, and instead, notice why he *isn't* -- despite having a brain which seems capable of processing information and making decisions. What might happen if he *did* "just" accept the pain? What about from *his* perspective? Might he know something we don't?

He actually tells us why² he can't accept it. Did you catch it?

"That's honestly the worst part. If I was speeding or something and crashed, like you said, doing something stupid, then I could accept it."

If it was just from him doing something stupid, then he could accept it. It wasn't, therefore...

Accepting that he could do "everything right" and *still* end up whacked like that is a terrifying thought. [There is no safe defense, not even science](#). Until he learns this -- which is not at all obvious, hence the need for the post -- then going on with his life like "Oh well, shit happens, nothing to see here!" would be missing something very important -- and setting himself up for being blindsided again. It seems pretty reasonable to want to avoid that.

It doesn't seem like much, and it wasn't much, but putting to words this "worst part" and validating the idea that "Yeah, that does suck, and it's actually not your fault, and here's what you can do to try to deal with this newfound aspect of reality", is actually active ingredient here. We're confirming "No, you're not missing anything, yes, this really is how it is, and here's how to get started orienting to it" -- and once he knows that, he no longer has that reason to postpone accepting the situation for what it is. By negating that reason, we've already taken a big chunk out of the problem by the time he's finished explaining what the problem *is*.³

He says a few other things in this section of the transcript which give clues to where he's coming from and why this might be a tricky situation for him to navigate. Don't worry if you can't articulate exactly what these things tell us, just pay attention to the sense you get when reading over the transcript. Where would *you* have to be coming from in order to say these things. Do you notice a few lines jumping out as "important", in some way?

There are multiple things going on here and it's not realistic to expect to see them all in advance (I didn't), let alone plan a step by step intervention, so don't worry about getting everything. What *can* you see? What *might* be there? Or if you don't think you see what the answers are, what *questions* might you ask, in order to figure out what is happening?

Next week, I'll share the rest of the transcript, along with annotations showing where I was coming from and how I chose what to say when.

Notes

1. What we refer to as "intellectually knowing" differs from what we call "alieving" in meta level.

When you're afraid of ghosts and "believe ghosts aren't real", you might say that you "intellectually know" that ghost aren't real which is to say that you think the belief that "Ghosts are real" is *false* -- and perhaps silly.

When you say you "alieve" in ghosts, that's another way of saying that you sure *act* like you believe ghosts are real -- and dangerous -- but you don't want to say that you *believe* ghosts are real and dangerous, because that would be stupid. And I mean... we can't believe anything *stupid*, right? That would make us stupid!

We don't actually need two different terms once we recognize that 1) sometimes we actually believe stupid things, and that 2) "intellectually knowing" a thing refers to a belief about what beliefs are *true* -- and that this is distinct from believing the thing.

Later in this sequence we'll go through an example of how keeping track of this meta level distinction and noticing when we're believing seemingly stupid things can resolve "irrational" fears. ↩

2. Well, one of the reasons why. ↩

3. The rest of this conversation takes longer in significant part because I wasn't able to immediately intuit what the other big block was. ↩

Chronic pain challenge part 2: the solution

In the [last post](#), I introduced an example where I was talking to someone suffering from chronic pain. Today we will dive into the rest of the transcript. Interspersed are annotations describing my thought process as the interaction unfolded. I cover what I did which seemed to work, and what I did that a cringe at in hindsight. Ten thousand foot overview at the bottom, for those who don't want to wade through the whole thing.

This might be a little difficult to follow, at this point. Originally, I was planning on posting this at the end of the sequence so that you would have the entire 17 posts worth of explanation to put my moves here into context and help understand where they came from and what I mean when I say jargony things like "bid for attention". You might want to revisit at the end of the sequence, and see if you pick up on more of what is going on.

I decided to move it to move it to the front for two reasons. One is that it can help to be primed with examples of the thing to be explained so that you more easily recognize the explanation when it fits. The other is just to demonstrate that when I say things that are very counterintuitive and difficult to believe, *there's reason to believe it*. It actually works.

Picking up where we left off...

other-guy: I suppose I have time now to consider what career path I want to take now and other hobbies to get into.

Jimmy: Assuming you can get your nerves and shit working again.

Up to this point I have just been indulging in my curiosity. What's going on here? What kind of hurt is he, and is his relationship to pain actually causing him problems or is it just injury that I can't really help with?

Often, the way to help people complaining about "pain" is to understand that it's not about the pain, and address the injury (forget the finger, look at the moon to which it points). Accept the pain's bid for attention, give it, and watch their suffering ease as they follow). My cousin's burn is an example of this, where people were trying to fight the pain because in *their* minds they had done everything needed to address the injury, and all I needed to do was make sure that in his

mind that was true too. The generalization of this that I've learned is that generally it makes more sense to stfu about your theories how things work and just focus on being the best person you can be. Care about your friends and family, try to solve their real world problems, and the "psychological" follow.

I didn't have to tell my cousin my stance that pain doesn't feel like an issue once you've listened to it, I just offered to listen with him. I didn't have to explain CBT to makeup girl and that she shouldn't take irrational feelings seriously, I just realized that her insecurity wasn't rational, and didn't take it seriously. By expecting a lack of suffering, suffering becomes surprising, and you get curious about what's causing it. By the time you walk that road to the end with them, your maps of reality have been updated together so that their expectations and reality can meet again, which is another way of saying "there's no more suffering". So usually I just do that.

This one was a little different. After getting the basics, I had mostly just run out of curiosity. I wasn't surprised that he was suffering, and I kinda felt like "I get it. And yeah, that sucks". That's not to say that continuing down the road of examining what exactly is causing his suffering would not have worked, just that I wasn't curious so I'd have to fake it and force it. Asking questions when you aren't actually curious what the answer is, because you think you "should" because it will "help" with the psychological problem, is failing to notice the hint from reality that *this information isn't important*. This is not the way. That's what a lack of curiosity means; it isn't important to you, so far as you can tell.

Since I was out of curiosity, and still wanted to see him get relief, that must mean I think I have something to say. But what do I have to say? It's not "You've integrated this information already, and can safely move on" -- because how the heck do I know. And I can't switch back to "Oh, I should ask, because how the heck is he supposed to know? "Have you learned your lesson?" and "Have you done what you need to do?" aren't particularly productive questions because what the heck did he do wrong? Allow the doctors to treat a condition he had? That's hardly a dumb thing to do, and what are you gonna do, *not* get the treatment you need? And what the heck is he supposed to do? He *tried* going to the doctor, and it didn't work, and there isn't exactly an obvious next step that he's avoiding. So far as I could tell, the man had done everything he could see or reasonably be expected to see, and was just trying to get the pain to *shut the fuck up* for a while so he could rest. Which is completely understandable and completely reasonable.

So I can't "Just be curious, and care about his suffering" because it was too predictable. Can't "Just tell him that he's okay and did everything he needed to" because I don't know that. Can't ask him if he has, because I already know he thinks he has -- and for good reason. I also have a *pretty good guess* that he's wrong -- which is something it's probably important for him to know, and something I should therefore explain.

Okay, great, so it's my turn to talk. And the thing I need to convey is that he needs to actually listen to the pain and take his problem more seriously, and that he's wrong in thinking he already done that. How do you expect simply saying *that* to go?

Probably similarly to what I anticipated. It wouldn't make sense. He wouldn't believe me, because he had no place to put such concepts. He might keep the idea "I'm *supposed* to believe this", which might come out as "I believe you", but it wouldn't be a belief -- it wouldn't be what his brain thinks reality *actually is*, and therefore wouldn't be the experience his brain constructs for him. Hypnosis is all about getting pieces considered even when they don't fit, so going hypnotic is one option. In context, it would probably mean explaining that the thing I was going to tell him is going to seem *very* weird, and leaning hard on the importance of suspending disbelief -- which, to a degree I did end up doing a little.

More importantly though, why work so hard to create room for a thing that doesn't fit, which is of limited use when it doesn't fit because he wouldn't even know how to use it to navigate, when I can just *explain* and *create* a way for it to fit? That way I don't have to lean so hard on "faith in the apparently absurd" and he'd actually know what to do with it.

So that's what I aimed to do, and that's *why*, in this case, it did come out looking a lot like "explaining the framework" rather than simply "doing the framework". It just happened to be that in this case, doing the framework involved teaching the framework, because the problem that I was interested in solving wasn't "this guy is suffering" but rather "this guy doesn't know how to navigate his suffering, and is therefore stuck".

So here's what I said.

Jimmy: I should give some more context about why I'm interested in what's going on with your issues there, and in particular the pain.

Jimmy: I've spent some time studying hypnosis and what that actually tells us about our brains and how to get what we want out of them. Pain is an especially interesting topic because it is simple, which means that failures aren't as muddled by "complex problems are complex."

Rather than simply asserting things at him without justification for why he should care or what my motives are, I'm framing it as explaining where I am coming from, so he can better understand what is interesting to me about it. I frame it this way to him because that's the frame I was operating from. That's just what's true, so far as I can tell.

I don't think he has to stay suffering, and I notice that he is, and I *am* curious about what it takes so that he ends up navigating out of this stuckness -- which is another way of saying I actually intend to help, if I can, and learn why I can't if I can't.

So I continue on with this explanation of where I'm coming from.

Jimmy: Hypnosis is pretty effective for pain, but what I've found is that it's not really necessary, and the—ugh, this is getting hard to explain, gimme a sec.

Jimmy: I'll suppose I'll start with an example. Some of my extended family and I were camping out in the mountains, and I went off to hunt for a bit. When I came back to camp, I found my little cousin sitting by the fire, holding his hand, and clearly fighting back tears.

Jimmy: Someone told me he had been playing with a piece of rebar in the fire, charring his initials into the log benches we had on our property out there, and that he inadvertently picked up the wrong side one of the times.

Jimmy: When I walked up to him, I didn't really know what to expect. I didn't know if he needed any medical attention or what it might be. I didn't know if he was worried that he would, or if he was going to be able to get whatever treatment it might need, or anything else.

Jimmy: My guess, based on the situation, was that it was just that it hurts. So I walked up to him, squatted down to his level as he was sitting on the bench, looked him in the eye, and asked, "Hurts?"

Jimmy: His eyes lit up immediately, and he replied, "Yes!" as if someone finally got it, finally saw what he was going through, only to immediately cut himself off and say, "Actually, no."

Jimmy: "When my brain feels pain, it interprets it as ticklish. It just tickles."

Jimmy: And just like that, the kid was happy and smiling, no tears in sight—just didn't hurt.

Jimmy: Which is a bit of a mind fuck, and people there had a hard time even noticing that he wasn't hurting anymore because how the fuck could he not be? Did you see his hand?

I feel a little embarrassed reusing the same stories, and I will give a more direct explanation later. But it's a good one, and goes to demonstrate the extreme difference between the boring reality and the seeming impossibility (before we notice how boringly normal it is). I start here

because "Pain actually isn't a problem, and *can't* be" is a super important part of my perspective that informs my curiosity and therefore must be conveyed.

It does conflict with how we normally see pain though, so I try to bridge that gap by explaining how it's actually congruent with how we normally treat pain without necessarily realizing it.

Jimmy: But we all feel pain that doesn't bother us, and the context and how we take it makes all the difference in whether it registers as a bad thing or whether we don't care or even maybe like it. Spicy food, massages that "hurt in a good way," etc.

Jimmy: But it applies everywhere.

Jimmy: My friend who went through natural childbirth said the pain didn't bother her at first, but at a certain point, it got hard to bear.

Jimmy: I asked her what that point was, and she said, "Oh... yeah, the point where complications came up and I was no longer comfortable in the knowledge that it was all going to plan."

Jimmy: My kid cousin had the benefit of growing up with a couple of older brothers (and a lot of older cousins), so he got "beat up" a lot in a loving and playful—but still rough—sort of way.

Jimmy: "When my brain feels pain, it makes it ticklish" was in reference to this rough play where it was tickling, and painful, and his brain had just noticed that even things that are quite painful can be safe.

Jimmy: And when things are safe, it's okay to relax and rest and enjoy the play.

Jimmy: His brain had made that connection in a context where "safe non-concerning pain" was essentially "tickling," hence the pain becoming "ticklish" when he noticed.

Jimmy: The reason he hadn't noticed earlier is that no one else there had given him space to. It was all, "Oh man, that must hurt!" "Get this kid some ice, because he needs it!" "When I burned myself like that, I was suffering all through the night and couldn't sleep."

Jimmy: This all told his brain that he was expected to be "not okay" and to suffer, that it wasn't appropriate to laugh and smile and play in times like these, that these couldn't be "mere tickles."

Jimmy: What I did was ask, "What is the problem, exactly? Is it that it hurts?" And when asked, and given space to find the answer to be, "Nothing, I guess," his answer was, "Nothing, I guess. Pain isn't a problem; it just tickles."

Jimmy: Our culture can be pretty bad at helping people in pain find ways of living with it—can actually pressure people into continued and unnecessary suffering, unwittingly—and I find that really interesting.

Jimmy: Like, what are the limits? My cousin's hand was all blistered up, and that didn't bother him.

The natural childbirth example serves to illustrate that the same principles apply even in supposed "extremes", and in natural cases with someone who hasn't been "hypnotized" or had any other "intervention" tried -- just a mentally resilient woman who is preferring to face reality so that she could better mitigate actual damage to her body. I then went on to explain that a lot of what leads pain to cause suffering is anxiety over not knowing what to do about it, and that the way to not suffer is simply to listen to the pain.

I really wanted to avoid coming off like a delusional "all problems solved with one simple trick!" weirdo, so I made sure to emphasize that I actually do recognize the apparent difficulty and am not flinching away from reality here. It's an *open question* to me, which I demonstrate by sharing some of the bits which have informed my perception when facing this open question.

Jimmy: One time I sprained my foot so bad my radiologist aunt was convinced it was broken—and that went from "so painful I can't focus on work" to "completely and utterly non-bothersome. Like the feeling of one's tongue in one's mouth; only in awareness when you're explicitly thinking about it, and unintrusively in the background otherwise."

Jimmy: That was the big eye-opener for me when I tried a certain changework protocol to address the pain (to my wife's credit, she's the one that called me out and said, "You claim to be able to help people with this stuff, but you're clearly suffering. Do something about it").

This is the "[acknowledgement routine](#)", and I put a description of what it actually entailed in this case into this footnote.¹

Jimmy: My friend with her childbirth fell off due to anxiety issues, but it wasn't an issue for her before that, and she actually wanted the pain because it helped her know when she needed to slow down to avoid tearing.

Jimmy: My wife's situation with our first was similar, where pain wasn't an issue until... I dunno, 24 hours of being in the hospital with complications, when fatigue set in and she no longer had the mental resources to contextualize the pain and hold it in what felt like the proper place.

Jimmy: I've encountered some situations like those where pain drugs have become helpful and useful, and situations where pain can be overwhelming and hard to orient to—partially intentionally, partly not.

Jimmy: But I haven't really found a limit where pain itself is an insurmountable problem or one that can't be addressed by disentangling the mental blocks that are preserving the discordance we call "suffering."

Jimmy: I also haven't had a chance to untangle chronic pain yet, or anything so unusual as actual nerve damage. That's why I'm interested in what your experience with the pain is like, exactly.

Jimmy: I want to know what that experience is like, how you're thinking about it, what happens when you play with things on the edges, etc.

I gave the best examples I could think of to demonstrate the limitations (and therefore my awareness of the limitations) and in doing so also validated feelings he might have of "I'm not just a little bitch, I don't know this will work, and if it doesn't that doesn't mean it's my fault" -- which I expected him to have because *I* would have them and they're very reasonable given the situation. So this serves to preserve security, by giving him a way to make sense of a potential lack of results without having to push back against me.

Jimmy: Anyway, sorry for the barrage of text. I just felt I should explain, and I don't know how to do so any more concisely.

This is probably unnecessary, but I did want to err on the side of being delicate about what might be a sensitive situation, especially because 1) he seemed sensitive about overwhelming me with long stories about this stuff if I might not be interested, and 2) knowing that I'm about to turn sadist on him and need him to feel safe facing overwhelming pain without flinching. Again,

mostly addressing security concerns here by conspicuously showing willingness to not threaten too much.

Jimmy: I guess I feel like... if I try to put myself in what I imagine to be your shoes... It certainly feels scary. And probably overwhelming both emotionally and physically, and I imagine I'd just want it to stop.

Jimmy: I could see feeling drained, and not-okay, and not really up to "playing" with the sensations that clearly aren't a joke and aren't the result of harmless play with big bro.

This is "pacing", to use hypnotist lingo. This is preemptively addressing "Am I missing anything?" so that when it's time to bid for attention, he has fewer objections to voice and more assurance that if his remaining objections were important I'd have noticed them and have taken them into account.

It's also more "security" stuff, working to attenuate any perceived pressure to face feelings he doesn't want to face. I anticipated (correctly) that this might be an issue, because we're so used to being bossed around by others and ourselves about what we're "supposed to" do, and accepting it. "Here's what you need to do to fix your pain problem" is something this all easily rounds to, but is very different and wouldn't actually help.

The fact that he *doesn't* need to do anything is actually an important active ingredient. He can do whatever he wants. That's what security is. That's what play is. It's what allows you to actually look at reality and figure out what you *want* to do given the situation. If you have to follow pre-prescribed rules either way then there's no place to look at reality and decide anything differently -- but it is exactly that decision that actually provides the update that does the work here. So if I'm not going full hypnotic and saying "here are your new beliefs", then it's important that he doesn't feel *compelled* to do anything.

Jimmy: I also really wonder what would be the resulting experience if I were to take a moment—or an hour, or a few days/weeks—to like... "recenter," I guess.

Jimmy: Like, on the "moment" timescale, "<Deep sigh>. Fuck. I guess this is my situation," and like... tune into the feelings a little bit. Feel them intentionally and really try to get a sense of the texture of the experience. What it's like to just be in that pain. And then take the next dose of drugs 30 seconds later than I might have, or whatever.

Here's my "payload" curiosity. This is the question I *am* curious on his answer to, now that I've given the context to understand it, and potentially find a real solution -- coming from his *actual embodied beliefs* -- rather than reciting what he thinks the answer is [supposed to be](#).

Jimmy: I guess I wasn't done rambling yet. It's just really interesting to me. Thanks for indulging me.

Jimmy: I guess the question I have—which I don't expect you to have an answer to, to be clear, but maybe we could find one—is, "What would it take for your brain to categorize the pain as 'just more sensations,' no longer a sign of 'something bad,' because the bad has already occurred and been factored in, such that you can move on with your life to the extent possible with the current physiological damage without also being hindered by no-longer-functional warnings of something bad distracting you and getting in the way of doing what you could be doing and want to be doing?"

Jimmy: Okay, I think I'm done now.

More "walking on eggshells" for security until I get the green light, restating the payload question in different words to try to hit the same thing from a slightly different angle. On reread, those sentences might be a bit hard to follow. Oh well.

other-guy: I've noticed something like that with my niece. When I'm watching her and she has a fall or something, I usually don't react like "gasp, OMG." I must be evil, 'cause I'll usually laugh, but she won't cry.

other-guy: When her parents are here, though, and they worry and gasp and ask if she's okay, she cries. I wondered if it's just 'cause she wants attention, and she knows I don't give it easily. Like I'll play with her and stuff, attention like that, but I don't hold her often and stuff like that.

Good, he recognizes the thing I'm pointing at. He may not understand all the nooks and crannies yet, but he's found the thing and has an existence proof.

other-guy: What you're saying makes sense. They overreact, in a sense, which triggers her brain to think it's supposed to hurt. But since I don't and I laugh, her brain thinks everything is okay and doesn't feel anything. It at least doesn't interpret it as pain.

other-guy: I'll have to try messing with it. My pain is like a burning pain, road rash, it's a cutting kind of pain, it's a pins and needles pain, it's like every kind of pain all at the same time.

"Try" is a red flag word. "Try" proves that they're tracking the distinction between intending to do a thing and succeeding in doing the thing, which shows they have some expectation of failure. You don't say "I'll try to get myself a glass of water" when you know you can do it, so he's showing here that he has reservations. This can be a passive problem in that he just doesn't get it yet, or it can be an active problem if the expectation is actually pulling him away from success.

other-guy: I'm wondering if my brain doesn't know how to interpret it, so it feels like everything constantly. The worst is when I shower. Heat and cold on my back sends me into orbit; showering really sucks. Interestingly, though, I hurt more after eating. Usually increases the burning feeling everywhere.

other-guy: Walking, after 5 minutes or so, increases the deep muscle pain, almost like it's my bones that hurt but similar to if you're squatting and your legs burn and you're trying to get one more rep in. Eventually, my legs give out. I literally couldn't move my legs anymore at the store and slowly started to fall. I had to have someone grab the electrical cart, and I managed to crawl into that.

other-guy: I went from doing 500+ pushups a day to barely being able to lift a gallon of milk overnight and lost about 40 lbs of muscle in like 3 weeks, like something was eating the muscle.

other-guy: The pain that shoots from my shoulder to my hand, at one point I even described it as if it feels like my hand is dying. It's really hard to explain that pain.

other-guy: This is really interesting. I've sat in silence just focusing on the pain before, trying to convince myself "pain don't hurt," like from the classic Road House, lol.

Yep, there's a problem. "Trying to convince myself" is ineffective, because if you're not immediately buying what you're selling, you need to be listening to why not. "Trying" just proves to yourself that you're not worth your own trust.

other-guy: I wouldn't say it helped, but it really made me think about my situation and how I just need to overcome it. The day it happened, in the hospital, they gave me so many drugs they couldn't believe I was even conscious, and the pain was still there.

"I wouldn't say it helped" shows that he's going to lose respect (in the meaningful sense of actually valuing my perspective -- I don't think he'd be "rude" enough to confront me on it) if I don't manage to show how this is congruent with my perspective actually becoming useful to him. So I gotta keep that in mind.

I get the impression that he doesn't expect what I'm saying to be of any help, but he doesn't want to sound dismissive or unappreciative of the effort.

other-guy: They eventually gave me ketamine, and that didn't take the pain away, but it made me not care. I remember during it still feeling the pain but just not caring about it, so I stopped grunting and screaming for a bit and got in a nap for a few minutes until they woke me up again.

Jimmy: Haha, doesn't it feel like you're trying to gaslight yourself with the "this doesn't actually hurt"?

Jimmy: I'm just imagining what that would be like coming from a doctor: "This doesn't actually hurt, what you're experiencing isn't pain."

This bit was a little forced. It is amusing, but a little too predictable to be as genuinely humorous as I tried to play it off as. This is an unforced error on my part. Oops.

*other-guy: That reminds me of my neighbor. One of the toughest people I've ever known.
[...]*

Jimmy: I'll get back to this when I have the spare brainpower to give a worthwhile response.

This isn't just a casual chit chat thing here. It's one thing to "do techniques" where you can rely on canned answers and "recite a script", but I'm trying to actually figure out what the truth is on many layers/objects, and therefore what is the action at each step that will bring about the best expected outcome -- so it's cognitively demanding. If you've been through similar cases many times before then you get good priors to lean on and things get easier², but here it's new to me

and I actually care about getting results rather than a per hour paycheck, so it takes some thought.

Specifically, I have to figure out exactly where he must be coming from to say this, what it takes to get him to where he needs to be, and what I don't know yet. And it's not a one sentence description like "he's coming from a place where he thinks X", it's an ability to put myself in his shoes and simulate his likely responses to things I could say by thinking about what *I* would say if I were in what I estimate to be his shoes.

OCT 18

Jimmy: What you're noticing with your niece is exactly what I'm talking about. When her parents gasp and worry, this shows her that this is something to worry about.

Jimmy: When they ask if she's okay, it suggests that she might not be. But even if they were to try to say, "It doesn't actually hurt, you're fine," when they're coming from a place of being freaked out about seeing their kid hurt, they'd just be showing the kid that they can't handle facing it, so they have to try to pretend it isn't real.

Jimmy: Which isn't a super reassuring thing to see from your parents.

Jimmy: It goes well for you because you actually aren't freaked out by seeing her hurt. Whether you say, "Oh no, I think we need to take you to the hospital!" or "Nah, you're fine, kid," the message that comes through is "this is actually fine" because it doesn't look like you mean the former and it does look like you mean the latter.

Jimmy: You've been there, done that, and if it were your own knees, you'd be singing the same tune because your brain genuinely recognizes scraped knees as something that is okay and not some big deal that needs solutions we don't have.

Jimmy: Which is why it can be really hard to do this for yourself because you know what it's like for you, making "This doesn't hurt" an even harder sell.

Jimmy: Whatever is happening with you doesn't feel like "no big deal, everything is fine," and given your description of losing 40 lbs of muscle and falling over when shopping, that sounds completely fucking reasonable.

Jimmy: So "Doesn't hurt, even" is a non-starter, really. It's gonna feel like something bad happened because it sure seems like something bad happened.

Jimmy: This is why I generally don't recommend trying to get to any specific outcome, and rather just explore what's there.

Jimmy: If you were to "overcome" this, I don't even know what that would look like. Do you? Because I don't think it's "remembering to just road house it," like it was for my cousin in a much milder circumstance.

Jimmy: And even there, I didn't have to tell him how to feel, and neither did he; once there was space to ask the question, he just looked, and that's what he found.

Just trying to explain that it's not about telling it's about listening

Jimmy: I could be wrong—and feel free to disregard anything I say if it doesn't fit—but it doesn't seem to me that what you need is more encouragement to devalue the pain and push through with things. I get the impression that you do enough of that already and know what it's like to function under healthy levels of discomfort.

Three things going on here.

- 1) More caution on all levels, showing that I'm not trying to push my ideas but genuinely offer them to the extent that I think he'll find them true and useful.
- 2) Validating his mental strength and willingness to put up with suffering. Often men like this are willing to put up with pain and suffering essentially as long as people expect them to, and need to be told when it's okay to stop.
- 3) Pointing at the idea that he might be better off actually pushing less, and therefore listening more.

Jimmy: I'm just curious what a solution would even look like.

Jimmy: Last time I said I was curious, "What would it take for your brain to categorize the pain as 'just more sensations,' no longer a sign of 'something bad' because the bad has already occurred and been factored in, such that you can move on with your life to the extent possible with the current physiological damage without also being hindered by no-longer-functional warnings of something bad distracting you and getting in the way of doing what you could be doing and want to be doing."

Jimmy: I don't mean to suggest that the answer is "nothing" or "next to nothing." It usually isn't. When I sprained my foot, it was "acknowledgement of my wish to not be injured, allowing in the full degree of suck that actually being injured meant, and then coming up with a plan for what to do differently next time (nothing, in that case) and what

else I needed to do to take care of it this time (get an X-ray, stop running on it like an idiot, etc.).

Jimmy: Often, it's something along these lines, where it's some sort of increased attention to whatever the error signal is pointing at—even if we think "we know already."

Jimmy: Because it feels like "obviously I know I hurt my foot, what do you want," it's easy to get misled off the track and miss why your brain is still calling for your attention in the way that it is.

Jimmy: So the question isn't really about what you currently think from the outside looking at the part of your brain that is generating the pain alarms, but for the part of your brain that is actually generating the alarms.

Jimmy: If we were to speak with it, metaphorically, what would it say?

Jimmy: Or another way of bracketing what I'm gesturing at, if you were to "sleep on it" and wake up with an answer, what would that answer be?

Jimmy: Because that's the kind of place the interesting answers are going to come from. From rolling around in the back of your head until new connections are made and just kinda come out on their own, rather than attempting to reason things through from things you already know.

Jimmy: I'm not sure if that makes any sense.

Just trying again to accurately articulate the question I'm intending for him to face, and inviting him to tell me if I'm doing it poorly.

OCT 19

other-guy: Sorry, had to read it a couple of times to understand. So with me, my niece doesn't cry because I don't treat her like something is wrong.

Yes! "I had to read it a couple times to understand" is a *great* sign. It means they recognize this as a non-obvious thing that requires more extreme attention ("going into hypnosis", to an extent) to grasp, and that he expects that he might have been successful in understanding.

other-guy: Kind of like how you can unknowingly cut yourself and not feel it until you see it or see blood.

other-guy: In my case, legs stop moving and get weak, freaks me out, and IDK what is wrong or what's going on.

other-guy: If I were to talk to that part of my brain, I imagine it would just tell me something isn't right, something is off, and that pain is its way of telling me or getting my attention.

other-guy: Hurting more at night I think comes from relaxing and being able to think about it more, so my mind goes back to the pain rather than during the day when I have stuff to do and it's not focused on it as much.

other-guy: Maybe I hurt from [redacted] because I think, "I hope I can do this, I hope I don't hurt much," so I'm focused on that—focused on the problem I didn't have a solution for yet—and it triggers pain.

other-guy: I always thought I hurt more at night because of everything I did during the day—overdoing it in a way, like a workout.

other-guy: Perhaps I need to stay busier, stop resting so much. Maybe try to work out again, and as more days go by, the more I forget to think about it—much like a breakup, which I'm also kind of going through at the moment.

other-guy: The breakup hurts now, but the more time that passes, the less it hurts, the less I think about her.

OCT 22

Jimmy: Having to read a couple of times isn't a bad thing. What I'm trying to get to is just past the seams of perception, which necessarily makes it non-trivial to grasp.

Jimmy: Right, if your brain doesn't know there's a problem, it can't react like there's a problem.

Jimmy: And yeah, your legs not moving seems like a problem, and that shit's freaky.

Jimmy: Something isn't right. Something is off. The pain is telling you and begging you to attend to this.

Jimmy: I know what you mean about things hurting more at night—both physically and emotionally. At night there's no more distractions; just you and your thoughts and your feelings.

Jimmy: You could try to keep busy and keep yourself distracted. I know a guy who does basically that. Drinks a lot, falls asleep with the TV on every night. Trying to get away from the quiet where problems are laid bare.

Jimmy: There are even some techniques for PTSD (e.g., EMDR) that seem to rely on distraction to keep things bearable enough to proceed with the technique.

Basically confirming his understanding, and preemptively "Am I missing anything?" on the "distraction" bit so that I can highlight an alternative.

Jimmy: The thing I'm trying to point at is that you don't necessarily have to get away from it.

Jimmy: Sometimes—well, I think people often miss the fact that pain isn't a bad thing. It's a sign of bad things, yes. And it can be overwhelming to be blasted in the face with the firehose of bad news.

Jimmy: But there's also this place where it's like... okay, yeah, it hurts. A lot, lol. And yet there's room to be at peace simultaneously.

Jimmy: Like... I wish it weren't that my body is all fucked up. Of course I wish it weren't that my body is all fucked up. And here I am, with my body all fucked up. Of course it hurts.

Jimmy: I'm still figuring out what to do with it all. Still all the pain, almost like when you were on the ketamine and "didn't care," only instead of "I don't care," the weight of the caring is still there. It's just "I know."

Jimmy: There's no "other side" against which the pain can pull you apart, because you wish you weren't all kinds of fucked up and you're feeling out what's there. Sorta "surrendering to the pain" so that you can explore it, if you will.

This is another important "payload", this time pointing at a truth rather than a question: "Pain isn't a bad thing". After delivering the suggestion I go on to build more context to help make sense of it so that it's something he knows how to respond to it and also how to hold onto it.

Jimmy: I don't know if that's a place you're ready to approach, nor do I suggest that you should push yourself to if you're not. There are reasons we flinch away from such things.

I don't know for sure he's ready to accept this suggestion, so I'm explicitly giving him the choice. Furthermore, since I anticipate that this might be one of those decisions like "just don't be afraid" that feels outside of our control due to poor representation of our reasons to choose one of the potential options, I'm going out of my way to explain why maybe he *should* take the

option that seems obviously wrong. This preserves the ability to weigh the pros and make a reflectively endorsed decision.

Jimmy: If someone is picking on you, calling you fat, for example, it really takes the wind out of their sails to respond, "I know."

Jimmy: It's hard to keep making fun of someone who just keeps agreeing with you, like, "Haha, yeah, I'm pretty big. They even gave me my own zip code."

Jimmy: The cost is that it means affirming the perspective that you're a fat kid and it's okay to make fun of you.

Jimmy: And if you affirm that unthinkingly, now you've taken on the identity as "the fat kid" and have established a precedent of not pushing back.

Jimmy: Until you know how and when to reverse these decisions, it can make a lot of sense to refuse the labels, even if they have some truth.

Jimmy: It can be easy to miss that the possibility is there, though, and I'm wondering if that might be why you're seeming stuck.

Jimmy: Does it feel more like "It's just an overwhelming pile of emotions, and I'm not ready to deal with them all" or more like "I didn't realize there was an option, and I don't entirely see what the options even are"?

Jimmy: I'm sorry to hear you're going through a breakup too. That shit sucks.

Basically explaining the costs of acting secure without excess clarity, so that he can feel secure choosing insecurity rather than closing his eyes to that facet of reality and fighting against his own reasonable desires.

other-guy: The ketamine and "calling you fat" are great examples. It makes more sense that way.

other-guy: For me, it's like the latter: "I didn't realize there was an option, and I didn't entirely see what the options even are."

other-guy: The pain sucks and is constant, and I think I've accepted this new me, and this is my life now. Well, maybe 75% accepted it.

other-guy: The pain really just makes me mad and I haven't found a way to release that emotion, I guess, and always try to distract myself with TV or YT or [redacted] or whatever.

other-guy: Before I was even on anything for my RA, I could still work at the time, but the RA pain was pretty bad. To deal with that pain, I started drinking after work, then drinking more, then drinking before work, and it turned into a problem. But it was the only thing that helped the pain.

other-guy: Everyone thought I became an alcoholic, trying to get me help and didn't believe me that I'm not—it's just pain.

other-guy: I finally saw a doctor that gave me some meds and a steroid shot, and two days later, I stopped drinking because it didn't hurt so bad anymore.

other-guy: I can deal with pain usually, but when it hits a certain level and becomes constant, that's when I get mad and frustrated.

other-guy: The breakup kind of sucks—[>decade long] relationship gone. No attractive girl will want a cripple with no job, lol.

other-guy: Thank you so much! I can't tell you how much I appreciate you.

OCT 24

Jimmy: "For me it's like the latter: "I didn't realize there was an option, and I didn't entirely see what the options even are.""

Jimmy: Great. I thought that might be the answer. Now we can have some fun. I can be kinda a sadistic shit and enjoy talking people into pain ;)

This is what I expected. It might not be 100% true and missing here derails everything, so it's worth making explicit. Also, here's where I feel confident enough that I can push him into the experience of pain and that if I overstep it will be recoverable because it will make sense to him in context.

Jimmy: Sounds like you're not too happy with your new identity as "fat kid," and are kinda pissed at life for pushing you to accept it.

Jimmy: I'm gonna go out on a limb and predict that the reason you're pissed and haven't fully accepted it is that "accepting it" kinda feels like "accepting your fate."

Jimmy: You don't want the rest of your life to be this way. The time you already spent messed up is fine. Spending a few more years recovering is fine. Shitty, yes, but fine.

Jimmy: It's the idea that it's over, and all you got to look forward to is being a cripple in pain forever that pisses you off, and which you don't want to accept.

Jimmy: Am I wrong?

Jimmy: Because if that's the case, then yeah, fuck that. I wouldn't accept that shit either, and being pissed off sounds like approximately the right reaction.

Jimmy: It's not time to accept that which hasn't been determined.

Okay, I think this is the important piece. This is the piece of truth that he could sense but hadn't integrated which would have been lost "just accepting" the pain and feeling okay without doing something with the message. This is why he pushed back, rather than "just" treating pain like any other information. To him, the pain seemed to be saying "Your life is over. You're a cripple now", so if he says "okay", then it's "Okay, I'm a cripple", an expectation of being a cripple, and therefore no ability to work towards *not* being a cripple. At least, not with his heart in it.

This is also what would be lost if we had decided to "use hypnosis" to "not feel so much pain" -- or any other technique³ with that same Goodharted goal. It's a pretty significant thing to lose, which is one of the reasons I don't advocate for trying to use hypnosis, and is why I'm fairly opposed to "using technique" in general. It risks putting the cart in front of the horse.

Jimmy: I'm actually pretty optimistic. Enough to be legitimately excited to see how things might change for you.

This was (is!) genuine excitement. Positivity is great, so long as it's actually justified and not inadvertently teaching that you aren't optimistic enough to look at reality. The amusing bit about this optimism is that it can coexist with remarkably low expectations.

Jimmy: Again, correct me if I'm wrong, but it seems like you're at quite the low and not even hung up on the fact that you got there but just want to get shit better.

Jimmy: And the cool thing about that is that even situations like "in serious pain only half the day" start to sound exciting. Having twenty pounds less muscle than you used to starts to be a motivating picture.

Jimmy: I don't know what kind of recovery is possible, but I poked around a bit on the first page of Google Scholar results for "hypnosis rheumatoid arthritis" and found [this] (<https://sci-hub.se/https://www.tandfonline.com/doi/abs/10.1080/00029157.1975.10403784>). There were a few studies showing improvement too.

Jimmy: I don't even know where to begin looking for stuff on the nerve damage from the infusion, but the fact that you're feeling so much and describe it as "feels like everything" / "my brain doesn't seem to know what to feel" makes me wonder what would

happen if you were to go into the pain and begin to embrace the sensations as separate from thoughts of future limitations.

Jimmy: I obviously don't know what would happen, but I wouldn't be surprised by either a shift in perception or nerves actually finding their connections and healing. From what I understand about the place you find yourself, it seems like it would be exciting to find out.

To be clear, this is a declaration of maximum entropy priors. I have no idea what would happen and make zero claims. I've just been surprised enough times by weird physiological changes that I stopped expecting impossibility without evidence.

Jimmy: [Breakup stuff] is the only thing that tempers my excitement to push you towards the pain and "finding out."

Jimmy: You're going through a lot. This is a lot for anyone, even without the loss of a long-term relationship.

Jimmy: I could certainly see needing some time to process that kind of thing.

Just guarding against pressuring him into things he's not ready for again. Especially since I am excited about how it could go, and he probably doesn't want to disappoint me after I've put in work trying to help him. But it's not about me, so my excitement isn't a valid driver, and I'm trying to make sure that if he pushes himself it's because *he* is excited about the potential pay off.

And that's it. That's the intervention that did the work.

OCT 25

other-guy: You are 100% right about it all. I'm pissed off, mostly because it's not my fault. If I was speeding and crashed, I could accept it easier because that was my fault.

other-guy: It's interesting though. I spent the other night only focused on the pain, like you said "as if I could talk to it," and it was weird.

other-guy: Like I told it, "I know something is wrong, it's okay, I'm fixing it, I'm working on it," and the last two days, I haven't needed the pain meds as much.

other-guy: I even did a couple of pushups today. That increased the pain a bit, but I didn't mind it because I actually did some pushups finally.

other-guy: It's like I understood the pain and accepted its presence, as if we're both sharing this body, and we have to learn to get along for now.

other-guy: I know that's really weird. It wasn't easy, it took maybe two hours of me focusing on it, and it's not a huge improvement by any means, but it is to me because I feel a little more free, a little less stuck.

Jimmy: Hey, when was the last time you increased your pushup count that much in 2 hours of work?

Jimmy: Sounds like a worthwhile pay rate to me.

Jimmy: As to it sounding weird, you have no idea.

Which interestingly enough, he didn't immediately find very helpful. However, later here's how he spoke of the difference:

Jimmy: Still taking the pain drugs?

other-guy: Not nearly as much. I still take the Norco, but the crazy strong stuff is at most once a day right now instead of 3x a day.

Jimmy: Has it leveled off or still improving?

other-guy: It's leveled off. Hasn't gotten better since the night I spent 2 hours. I'm just happy it's not been so miserable.

Jimmy: Ah, that's good. Do you think that if you were to spend 2 more hours, you'd improve further?

Jimmy: In other words, is it that you feel like you're up against a wall on that front, or just that your motivation is elsewhere at the moment?

other-guy: It might. I've tried once and fell asleep lol. I haven't tried more mostly because I already feel more free and just been enjoying that, but I do need to try again.

Jimmy: Haha, okay, good. Just want to make sure you aren't stuck.

Jimmy: No need to rush these things. If you're enjoying only being in as much pain as you are, then enjoy it for as long as you need to.

Jimmy: Kinda funny how these things work.

Jimmy: I gotta admit I'm cracking up thinking about "I still take the Norco [and] the crazy strong stuff [up to] once a day" being relief, but I totally get it.

Okay, this one I was genuinely cracking up about. It still makes me smile to think about.

other-guy: Haha, right?? Like there's still a bit of pain, but it's a huge relief compared to what it was.

other-guy: Tonight or tomorrow night, I'm gonna try again. Spending another 2 hrs if I need to.

[...]

other-guy: Hey, my friend. Just a quick update. Still working out a little without pushing it. I tried another 2-hour thing and didn't really get better, but I still feel light-years ahead of where I was last month.

Jimmy: Cool.

Jimmy: I looked back over the chat, and the one thing I noticed is that there seems to be an inconsistency between how you describe things now and how you did right after the first 2 hours you put into it.

Jimmy: "It wasn't easy, it took maybe 2 hours of me focusing on it, and it's not a huge improvement by any means" vs "Like there's still a bit of pain, but it's a huge relief compared to what it was," and now "light-years ahead."

Jimmy: Which is interesting in contrast to "It's leveled off. Hasn't gotten better since the night I spent 2 hours."

Jimmy: Do the "not a huge improvement" words feel foreign now? Or do you have a good feel for why you were describing it in a way that seems different then despite also describing it as "hasn't gotten better"?

Jimmy: (This stuff can be weird. I think I mentioned how confused I was when I injured my foot and couldn't figure out when it got better or how it did because it also felt the same.)

other-guy: I think before spending the 2 hours, the pain was so intense. When it got better but still with some pain, I didn't realize how much better at the time.

other-guy: Maybe I'm just reflecting, looking back at doing very little to small workouts every day. It felt like it was going to take forever to get to this point, and it almost happened overnight.

Jimmy: So you're saying that you think it did get better right away and level off, and it just took you a while to recognize how big the improvement was?

other-guy: Yeah, I've been trying to not push it but trying to do more and more and still not hurt like hell is showing me how much better it actually got.

Jimmy: Is it a delayed thing, where if you do too many pushups you're feeling it later more than during the pushups?

Jimmy: And is it still a weird hard-to-understand "feels like everything" sort of pain? Or something more understandable?

other-guy: Mostly. If I do too many pushups, I start to feel it within a half hour or so.

other-guy: Otherwise, pain is barely noticeable while I'm active. Hurts more when I start to rest, but it's bearable.

other-guy: When it hurts, it is still more like "feels like everything."

other-guy: My apologies, man. I just looked back at the messages earlier and just barely understood the envelope part. Thank you again and again. I can't begin to tell you how much I appreciate everything, especially helping the pain.

other-guy: You helped more with a couple messages than doctors have in over a year.

After a bit more chatting,

other-guy: I honestly thought I had at least 2 more years of this before I could be at this stage. This level of improvement already is amazing. Thank you.

Interestingly, that intervention proved to be quite helpful to him despite not feeling like much at first. It reminds me of my own realization with the hurt foot, just on a larger scale. It can be weird to feel the same pain, and also really not want it to be there at time and be motivated to do something about it, and also not be suffering in the same way. We're not used to having to treat pain/suffering/meta-suffering as distinct things, so it takes some time to figure out what's going on.

He still doesn't have perfect solutions to everything, but it actually doesn't make sense to solve everything perfectly, and at some point you gotta go live your life. This is where we left it. If he wants a better solution I can help him find it, and he's just happy to have the solution he has in the meantime. My ego would definitely prefer a more complete solution, but also that's not a

good reason, and trying to act on such things isn't a thing that leads to good results in the most meaningful sense.

10,000 foot overview

The way it looked to me is like this:

He didn't know that this kind of suffering could be resolved, so he didn't know that he had something to figure out. So of course he didn't put much work into figuring it out. If he understood this, he'd be motivated to figure out how.

He didn't understand the structure of this suffering, so he didn't know where to look for a resolution. If he understood that the place to look is where the pain is pointing, he'd be motivated to look.

I didn't know if that motivation would be enough on its own, or if his object-level reason for not looking was going to need addressing, but once he had that motivation any obstacle that needed addressing would become more clear because it'd be in the open.

This all turned out to be more or less correct, which is why it worked out the way it did. It did turn out that there were important reasons for him to not simply yield to the pain, which wasn't surprising and was factored in to the approach I took. In hindsight, I feel like I probably should have seen those reasons coming, but hey, what're ya gonna do.

My first response did hit the nail on the head, and in response he explicitly told me that it's why he can't accept it. I immediately pivoted to giving attention to the reasons he couldn't accept it, obsoleting part of the reason he couldn't accept it. I just didn't see the rest of it clearly enough to continue this approach until the whole thing dissolved without intentional work towards dissolving.

In both this case and the case of my kid cousin with the fire poker, the suffering was resolved with a clear picture of what was going on. With my cousin, all he had to notice is that there was actually no problem.

Chronic pain guy had more things to notice:

"Suffering can be resolved by working through what's going on" -> "This probably means listening to the pain" -> "There are actually good reasons to not listen to the pain" -> "His particular reason is that he thought listening to the pain means accepting that he's crippled forever" -> "He doesn't actually have to accept that in order to listen to the pain".

Once he saw that, there was no longer anything blocking the default path of attending to information that was important to him.

UPDATE: I talked to him some more in September 2025, so just under year later:

I've cut way back on the pain meds, I really only take them for these intense migraines I started to have like once a week. Usually not for the regular pain though.

Notes

1. In that case, the chain of wishes/explanations went like this "I wish it wasn't that it hurts" (this one felt like "just words"), "It hurts because I broke my foot", "I wish it wasn't that I broke my foot" (this one felt *meaningful*, which immediately stood out as interesting. "Do I not actually care about the pain!?!"), "I broke my foot because I was playing football and shit happens". At that point, the "do you wish it weren't that <cause>?" was no longer "yes". I don't wish that I didn't play football. I like playing and having fun with my family, and a broken foot is a fine price to pay for all the physical play of this nature I have done. So no, I do not wish I didn't play football. Similarly, I did not wish that shit didn't happen. That felt like too ridiculous a thing to wish for.

So of course shit happens, and of course I played football. So of course I broke my foot. And of course it hurts. What do you *want*? Nothing? I still feel the pain. But I don't wish I didn't? It just feels *okay* now? Somehow? *Weird.* ↩

2. This is one way of conceptualizing "technique"; technique is the prior belief that these responses will be appropriate. ↩
3. This is also what is lost when you rely on pain killers. ↩

Part: Low-Level Mechanics of Changing Minds

The next three get into the low level mechanics of changing minds optimally, through direction of attention.

Expectation = intention = setpoint

When I was first learning about hypnosis, one of the things that was very confusing to me is how "expectations" relate to "intent". Some hypnotists would say "All suggestion is about expectation; if they expect to have an experience they will", and frame their inductions in terms of expectation (e.g. "Your eyelids will become heavy"). The problem with this is that "I don't think it's gonna work". Other hypnotists would avoid this issue entirely by saying "I *don't care* if you think it will work. Follow my instructions, and you will get the results regardless of what you believe" and then say things like "*Make* your eyelids heavy". The problem with this is that "I don't know to do that!", which would be avoided by saying "You don't have to 'do' anything; I'm telling you what *is going to happen*, and your job is simply to notice when it does" -- back to square one. This dual path thing always confused me, because "Sure, we can equivocate and play sleight of mouth in order to trick people into getting the results¹, but which is it *really*?".

It's both.

They're the same thing.

In the same way that "running away from a tiger" and "running to catch a bus" are both the same object level behavior being used for a somewhat different purpose.

Initially, it seemed weird to me that humans are designed so that expectation and intent are implemented in the same way. Strange design choice, given that it creates massive vulnerabilities, where if we can be tricked into expecting we won't be able to remember something, it becomes self fulfilling.²

Turns out, it has to be this way - not just for humans, but for any system that tries to control anything. To see why, let's start with something simpler than a human mind.

Consider a thermostat.

When anthropomorphizing a thermostat, you might be tempted to model this controller as having "beliefs" and "desires". The output of the thermometer is the "belief" the system has about the temperature of the room, and the set point we dial to is the "desire" of the system, to which it attempts to regulate external reality. Under this model, it kinda seems like intentions are different kinds of things from expectations, because expectations are like beliefs which are about what reality is while intentions are like desires which are about what we want it to be, but isn't. But a thermometer reading isn't really an expectation, so where's that come in? What does a thermostat "expect"?

Let's model with a bit more detail, including some necessary facts of reality that we skipped over at first. This is a little weird and confusing, so bear with me here.

Say we want to do *better* than a thermostat on the wall that either flicks the heater on or off depending on the most recent reading of the thermometer. That works "well enough", for simple systems and lax standards, but say we want to get it *right*. Say we want to put out the exact amount of heat to get the temperature *exactly* right, to the best it is possible to do.

Now, all of a sudden, we have a lot of complications to think about. To simplify it we'll still ignore most of them, but the following two are important.

1. Thermometer data are *noisy*. One of the better ways to measure temperature is with a thermistor, which is essentially a resistor whose resistance changes with temperature. But now we have thermal noise of electrons polluting our signal. How much noise is there relative to the signal? Well here's the equation for thermal noise:

$$\overline{V_n^2} = 4k_B T R \Delta f$$

If you try to calculate the noise for a single instant in time, that annoying little "delta f" term becomes infinite. There is no such thing as an instantaneous measure of temperature (or anything), because we need to accumulate signal over a finite window before it can stand out above the noise. The more time, the more accurately we can measure. Thermometer readings aren't *belief*. They're *data*, and can only approach belief in absence of disturbance.

2. Our own output disturbs the temperature!³ By design! We can't just take forever to measure our temperature and get perfect accuracy, because if we ever try to *change* the temperature we now have to do it infinitely slow or else we screw up our measurement!

When setting the AC thermostat, or checking the internal temperature of the turkey in the oven, these things don't really matter. The thermometer responds on the order of seconds, the turkey responds on the order of hours, so the predictable error here is "small enough" and we can ignore it. The moment you try to make an *optimal* controller of anything (maybe temperature control for a more sensitive chemical process in the face of external disturbance), this becomes a real issue.

For example, imagine we're trying to do optimal temperature control, and we get to cycle the relay on and off once every ten seconds. This means that every ten seconds we have to decide for how long we want to keep the heater on, and for how long to keep it off. At each time step, we have ten seconds of new thermometer data to update on. What do we do?

Bayes' Law, right? The most probable temperature is the one you get by starting with your priors -- i.e. your last temperature estimate, based on the previous estimate, new thermometer data, etc -- and update that with the latest thermometer data and whatever else you might know.

But now we have to decide how much to trust priors and how much to trust new measurements. "Look at the [new] data! Don't bias your results!" is one way of going about it, but to the extent that the new data is noisy it *makes sense* to look back at your priors built on old data, and average over more data. If we do that for long enough, and have no unpredictable disturbance, we can get perfect measurements! If there's a disturbance though (maybe the wind is blowing), then we're back to the issue of the actual temperature changing faster than we can measure it -- so who cares what temperature it was *yesterday*, we need to look at what the thermometer is saying *now*. If we have any disturbance at all, and any noise at all -- which we *always* do -- then we can never know the actual temperature perfectly; we can only calculate what it will be *in expectation*.

Once we have our temperature estimate, we can decide how much heat to add, and that messes with the temperature estimate again. We might expect, based on previous thermometer data and the latest batch, that the thermometer reading at the next timestep will be 60.0F. But if we turn the heater on full blast, it might increase the temperature 10 degrees, so do we expect 60.0F, or 70.0F, or somewhere in between? We can predict 60F and be right, or 70F and be right, or anywhere in between. It's underspecified, whatever shall we predict?

So long as you desire it to be somewhere between 60F and 70F, you predict it to be *exactly* what you want it to be. If you want to get to 65F, no more no less, then you add exactly 50% heat. This brings the temperature up 5 degrees from where it would have been, and shifts the expected temperature up 5 degrees. Not only do you expect that the temperature will be approximately 65F eventually in the future sometime, 65F becomes your *exact proximal expectation*. For if you expected any less at your next measurement, you would output more heat, and if you expected any higher you would output less.

You *don't know* what the exact temperature is at any moment in time, and the exact temperature depends on what you're putting out. If you last got to read the thermometer five seconds ago and it said 64.5F, do you think you're a half degree low? No! Because you put out enough heat that you expect that to be compensated for, so you have *no idea* whether you're still below 65F or not. The act of optimally controlling to a set point -- which could usefully be termed "actually trying to reach the setpoint" -- *sets* the expectation of what external reality is to be equal to your setpoint. To the extent you can be said to have any intention whatsoever, your expectation is *necessarily* the same thing as your intention, or else it's not a real intention. Because you aren't doing what you need to do in order to realize that desire, even in expectation.⁴

And this is why expectation is equivalent to intent.

You have a variety of output behaviors you can choose from, and with them the corresponding set of expectations you can have. Choosing what you expect to happen goes hand in hand with choosing to act so as to realize the corresponding expectation, and this is what intent is; deliberately acting as to expect that result. We can expect without *deliberate* intention, but we can't intend without expectation, and regardless of how deliberate we are about setting our intentions, we will always behave in ways that realize our expectations -- in expectation, by definition.

Returning to our simpler thermostat model (or our turkey model), where the thermometer really does read out a good estimate of the system temperature which is predictably below the desired set point for minutes at end, this is the result of *limited control*. Here, we *can't* expect the temperature to be equal to the set point -- without closing our eyes to the data and refusing to take it into account. We can expect it to *become* equal, but no turn of the dial will cook a turkey as fast as we can orient to incoming data.

If I blast full power to my homemade heat treat oven, the heating elements will melt in seconds. As a result, I programmed it to limit the output power to what I expect to be safe (even temperature controllers have insecurity issues), and it is better modeled as having an intention to control the temperature to *as close to desired as is safely possible* -- because that's what determines the heat output. It controls to a *trajectory*, which is the optimal trajectory given the limitations. But when I first turn it on, and it's 2000F too cold, it has no expectation of being at that temperature, no intention of being at that temperature -- only an intention of getting there, if it's safe.

My heat treat oven cannot, and we cannot, intend to do anything which we cannot expect. What we can do instead, is expect and intend to be on a trajectory that leads to the desired set point. This distinction between "where we want to end up" and "where we intend and expect to be *right now*", and at our next time step", is where the confusion comes from.

So what's this look like, as applied to humans? How does this expectation/intention equivalency and "end state vs trajectory" thing confuse people in practice?

Sleep talking, on purpose

For a slightly weird example, one time I tested this by asking my wife to remind me to turn the oven off when I got in bed, knowing she'd be asleep. Her initial response, quite understandably, was to object "I'm going to be asleep!"

I told her that I understood that, and that I wouldn't be upset if she failed, but I wanted her to promise me anyway. Go through the motions, make the promise, and if you don't do it I will forgive you. She said okay, she'll remind me, and went to sleep.

When I got into bed, she reminded me to turn the oven off, and in the morning she had no idea she did so. She had been asleep.

I could have dressed this up as a "hypnotic induction" and rambled about how "it's not her, but her *unconscious mind*" which would enact the suggestion -- using that to try to *trick* her into not refusing my request. But I could also just tell her that if she fails it's fine and let's find out if she can do it. It's still her, no need to "wooo wooo!" about it, and no need to deny reality in order to find the truth. Even if she was literally unconscious at the time of fulfilling that expectation.⁵

You can just... *decide* that?

For an even weirder example, imagine trying to "decide to not swell" an injury you have. It's so bizarre it feels like "Where would I even start!?", because it's not clear where those levers are and that makes it hard to expect it to work. Which necessarily makes it hard to intend for it to.

I met a hypnotist once who told me about how his friend had hypnotized him to help him with his elbow swelling, and I was super skeptical. I didn't write it off completely, but I sure didn't feel like that's something I could expect to work. As a result, I never intentionally tried, because I don't like telling people to expect things I couldn't expect myself, and I certainly couldn't expect it to work. That seems *crazy*.

And then one day at Jiu Jitsu, someone ankle locked me before I could tap, spraining my ankle pretty pretty badly. I was quite frustrated, and I really didn't want to have to deal with a swollen ankle keeping me from doing even things that *wouldn't* overstress a healing ankle. I found myself indignantly refusing to swell the injury. I have no idea where that came from. I did not feel like "I chose to do that", I felt like "You can't just *decide* these things you crazy person!!!"⁶. At the same time, I *didn't care*. "Don't care, doing it anyway. *It's not going to swell*. I'm not gonna do it. I just won't take advantage of it, I'll take care of my injury, and it doesn't *need* to swell". Okay, I thought. We'll see what happens, I guess.

It didn't swell. Took a whole month to heal, but never swelled up like my previous ankle sprains had.

The next time I expected an injury to swell came when I dropped a small boat on my finger, smashing it between the boat and the concrete at the water's edge. I *knew* from experience that it was definitely going to swell. It wasn't a borderline case. Even knowing "Hey, so, these things are apparently decidable" didn't make it feel any more decidable than you'd expect. It's like

telling someone "Oh, you don't have to keep an irrational fear around. You can just decide not to fear". Yeah, sure, buddy.

But I knew it was likely *true*, since my previous experience was compelling, so I put some work into it. Yes, it certainly *feels* like it's definitely going to swell up. That's certainly a reasonable expectation to have and it's backed by a lot of... confirmation bias, at least. I had a lot of experience confirming that jammed fingers would swell when I expected them to, but not a lot of experience with things swelling when I expected them *not* to. What if it didn't *have to* go this way? *What would it be like*, if somehow, it were to not swell? Could I imagine that? Could I imagine that *being real*?

With some effort... actually, yeah. I could imagine what that would be like, and given my previous experience I could imagine that actually being true. So I went with that, and sure enough it didn't swell; I could still bend my fingers just fine. Which was kinda a surreal experience too, given how smashed my finger got. Since then, it hasn't even felt like something I shouldn't have significant control over. I just think about what I want to do, and expect that.

I told my friend about this "Apparently you can just decide to not swell injuries" thing shortly after my first experience with it, and she was understandably skeptical. At the same time, she knew from experience that when I say something that sounds hard to believe it's because I have better reason to believe it's true than she does to think it's false -- and that these things tend to prove to be true. So she knew she couldn't actually rule it out, and so next time she got injured, she tried to expect it to not swell. And did expect it to not swell. And it didn't swell. All it took to give her this same ability to decide when to swell things was to casually mention that she already can, and tell her about the one experience I had.⁷

You [don't] have to believe!

You know how high school sports coaches like to go on about how "You have to *believe* you will win!"? And how the standard rationalist response is "Nonsense, of course you don't. Beliefs are supposed to track reality, not be wishful thinking. Believe what looks to be true, try your best, and *find out* if you win"?

The coach does have a point though, and there's a reason he's so adamant about what he's saying. If you *expect* to lose -- if you're *directing attention towards the experience of your upcoming loss* -- then you are intending to lose, and good luck winning if you aren't gonna even try. The problem is that he's expecting on the level of "Will we win this game?", which, according to the data, isn't looking like it's something we can control. He doesn't know what else to do, and he doesn't want to *just give up*, so of course he's going to engage in motivated thinking. Fudging the

data until he can expect success is the only way he can hope to succeed. It's a *load bearing* delusion.⁸⁹

One way to do better is to deliberately trade correctness of expectation for effort *without* letting delusion spread to infect the rest of your thinking. "Yeah, I'm probably going to lose. I don't care. I intend to win *anyway*". Or, in other words "Do or do not. There is no 'try'". That means setting yourself up for failure, expecting success *knowing* that you aren't likely to have that expectation realized. It's not pleasant, and that gap between your expectations and the data coming from reality *is* what suffering is. But with suffering comes hope, and sometimes the tradeoff is worthwhile.

The other option, which tends to be preferable when you have time to compute it, is to shift from trying to predict instant success to predicting that you will stay on an optimal trajectory. Instead of coding your temperature controller to try so hard that it melts down if there's ever a large gap between the belief and desire, limit the output power. Code it such that when too cold, it will heat *as fast as it safely can*, and predict that the temperature will therefore rise as quickly as is safely possible -- which may or may not reach the desired set point.

In other words, "Believe what looks to be true, try your best, and *find out* if you can win".

To give a concrete example, in the finals of the most important tournament of my high school sports career, I went in knowing it'd be a tough match and very quickly realized that I definitely wasn't going to win. Not "Wah wahh, boohoo, I'm not gonna win!", dramatizing a fear that I *might* not win, just a sober recognition that I got second place. Which hey, could be worse.

It's one thing to talk the talk of "Just try your best and find out what happens", but a true 100% effort is *hard*, and *risky*¹⁰, and what's the point if not to maybe succeed? Normally, I fought with an intent to win. My motivation that justified all the effort was in expectation of potentially securing a win, even if somewhat unlikely. This match started so badly that I quickly recognized that it wasn't gonna happen. I couldn't even imagine a scenario where I could win, and as a result I had no ability to control towards a success. Without that hopeful expectation of success I couldn't motivate a real effort. I couldn't *try*.

I grieved my loss for a moment, pondered for a bit, and ended up deciding that five more minutes of maximal effort was worthwhile just out of principle. I decided that that I would intend to do *as well as I possibly could*, giving up absolutely nothing that wasn't taken completely against my will. With this alternate intention, it *didn't matter* that I was definitely going to lose. I knew it was true, and so there was nothing there to see. No reason to look at "Might I win?", and therefore no reason to expect "No". And nothing keeping me from denying my opponent every point I possibly could while scoring every point on him that he let me. With

about thirty seconds left I realized "Oh shit, I might actually win", which was scary because that meant I might lose (which, amusingly enough, isn't a fear so long as you know it's reality).

Surprisingly enough I *didn't*. Turns out, "Denying every point possible" was all of them, and "Scoring every point I could" was more than he had scored already.¹¹

If I had expected on the level of "Will I win?", even attempting to fudge the scales in effort to predict and therefore aim towards "Yes", I would have predicted failure, behaved in line with that prediction, and never figured out that it didn't *have to be* true. Even my coach admitted after the match that he didn't have that optimism in him.

Motivated thinking often does this. We try to believe what we want to happen, and struggle against reality telling us we're failing, only to miss the chance to do something better. Give in completely, take your thumb off the scale, submit to the reality of what you cannot expect to change, and you get to notice the question of "What next? Given that I'm [probably] going to lose, what *do* I want?". While likely not enough to achieve everything you ever wanted (by your own estimation), more nuanced models that predict optimal trajectories tend to map reality better than those which predict immediate and unconditional success, and work better as a result.

Pay attention

To what you expect, because that is what you intend, and that is the world towards which you are aiming. Not to what you *would* expect, if you were to attend to a particular question -- to what you are actively expecting in each moment.

To the difficulty with which you can expect to succeed, for that is the warning sign that even trying is going to require submitting more to reality, and finding something you can expect less ham-fistedly -- something that is as good as you can possibly achieve, and therefore you might *actually* achieve.

Notes

1. I sat in on the last class of a hypnotherapy training course once, and the instructor (whom I respect) was saying that we are essentially *con men*, conning people into getting better. ↩
2. It's not *quite* self fulfilling, technically. No one is ever unable to remember their names in hypnosis. People are unable to *try* because they've been convinced they can't succeed -- generally without recognizing that they've been convinced and that the apparent inability stems from the lack of intent. The end result is functionally the same, until you figure out that you can stop imagining that any time you like -- [at which point the vulnerability is patched](#). ↩
3. Goodhart's law is a bitch ↩

4. This isn't a "No True Scotsman" because it literally wouldn't fit the definition of "intention" otherwise.

in·ten·tion: a thing intended; an aim or plan.

If we're actively updating our behavior to maintain expectation of something else, that something else is the thing we're aiming at. If we're updating our expectations and not changing our behavior in response, we're not aiming at anything.

↩

5. This is the opposite of hypnotic "challenge suggestions" like name amnesia or sticking hands to tables. ↩
6. Having ones inner monolog replaced with a legit dialog involving accusations of insanity does not help one feel sane. ↩
7. It did take a little more work to help her *keep* it, several injuries down the line.

Eventually, someone offered her ibuprofen when she broke her thumb, and she casually said that she didn't need it because she just decided it wouldn't swell. This was such an unlikely thing to him that his brain autocorrected over it, as if he had seen "[the the](#)" and reasoned that it couldn't possibly be what was said/meant. She felt self conscious about it, started doubting herself, and her injury swelled up. I had to talk to her that night about "Who cares if it sounds crazy? *Find out* if it's true", and the swelling was most of the way back down in the morning. ↩

8. It also applies to the [the makeup washing off example](#). Her friends wanted her to get into the pool, but couldn't look at reality and expect her to go swimming, so they didn't. In order to preserve their expectation that she goes swimming, they dissociated from her feedback and cut off any chance of anyone learning anything that could change their minds -- and because this is visible, it means their beliefs don't track reality and can't be used to [Aumann update](#).

This is where the apparent "other sense" of the word "expect" comes from -- people can and do choose to expect things (direct attention and attempt control towards outcomes) while simultaneously expecting to fail at realizing those expectations, leaving a sense of social pressure without necessarily expecting that you will yield to it -- just that you "should". As strange as it sounds, her friends were genuinely expecting her to get in the pool, while simultaneously expecting her to not do as they expected her to.

Yes, this is fundamentally irrational, and it's for a reason; they had no idea how to even *try* if they were to be rational, and it's better to choose a slim chance at success than it is to choose *no* chance of success -- what is rationality for if not for winning?

My thesis here is that no, full rationality really is better. Of course you don't know what to do with the discouraging facts right away; you have to figure that out! ↩

9. Another important corollary is that not only is motivated cognition fundamental, so is confirmation bias. When we teach kids that the scientific method is "come up with a hypothesis, and then test it", that's simply wrong, and a recipe for confirmation bias -- because by focusing on one predetermined outcome you are programming your brain to confirm it. This is fine as an *engineering* hypothesis where you test "I can make a cool thing work", but good science is going to come when you *don't know* what will happen when you do a certain thing, and run tests to *find out*. The "hypothesis", in good science, is "This experiment is going to teach me something useful!". ↩
10. Because of this match some of my ribs pop out further than they're supposed to, to this day. Oops.

More dramatically, Eddie Hall's record breaking [500kg lift](#) had [some scary effects](#) on him, and to get himself to output that level of effort [he had to feed his brain false data to get himself to treat it like a life or death issue](#). ↩

11. I didn't have an explicit understanding of how expectations shape intentions, or how to consciously navigate from using my attention in an ineffective way to one which was more effective. But I did have a stubborn rejection of thumbing the scale, and a stubborn rejection of the idea of *not trying*.

The reason this works as an example despite preceding explicit my explicit understanding is that it's not explicit understanding that is doing the work. We already make these decisions when we notice them, and our results follow our decisions. ↩

Setpoint = The experience we attend to

In the [last post](#), I talked about how hypnotists like to equivocate between expectation and intent in order to get people to believe what they want them to do and do what they want them to believe [sic].

There is a third framing that hypnotists also use at times, which is "imagination". You don't have to *expect* that your eyes will be too heavy to open, or *intend* them to be. You can just *imagine* that they will be, and so long as you stay in that imagination, that is enough.

The most simple version of this, which you're probably familiar with, is the phenomenon where if you imagine biting into a sour lemon your mouth will begin to water. Or perhaps more frequently, you might imagine other things and experience other physiological phenomena associated with that imagined experience. Either way the point is clear, our brains and bodies begin to respond to imagined experiences almost as if they're real. "Imagination", at its core, is simulated experience. It's trying ideas on for size, figuring out what that *would* be like, how we *would* feel, how we *would* respond -- and then hopefully inhibiting those behaviors eventually when we believe the situation we're simulating to not be "real".

If I ask you to "imagine biting into a lemon", you have one experience. If I ask you *what it feels like* to bite into a lemon -- assuming you don't lazily give me a cached answer of "sour" -- you simulate that experience and report on your experience. If I tell you that I'm *going to give you a lemon to bite into*, and you think "I love lemons! I'm gonna bite into a lemon!" and start thinking about *what you're going to do* (What you "expect" to do? What you "intend" to do? Kinda starting to feel the same, huh?) then it's the same thing *again*. Still the experience of biting into a lemon (while not *actually* biting into a lemon), mouth watering, just with a different framing of "imagining" vs "figuring out what it *would* be like" vs "thinking about what is *going* to happen". Same object level mental representation, different meta level tags.

These "meta level tags", for lack of a better term, do matter. It's what keeps us grounded. It keeps our mouth watering only a little, keeps us from being too mad at our spouses for what they did in our bad dreams, etc. But they matter only to the extent that it drags attention away from the object level thing. We're not perfect, and so sometimes people *do* get mad at people for what they do in dreams. And our mouths water a bit, and all that. And hypnotists can abuse this arbitrarily far.¹

On the object level, we're either attending to an object level experience of reality, or we're not. Unless we know we're "just imagining", or unless we're not actually experiencing it, we don't know the experience isn't *real* (another one of those stupid tautological statements). If we're

aiming our attention at "I will hit this target", then when the sights drift right, we shift our aim to the left -- because otherwise we'd quickly stop expecting to hit the target. If we're aiming to keep the room at 70F, then when it starts to feel like 69F, this new data is uncomfortable until it starts feeling like 70F. And again, we either act to alleviate this discomfort or otherwise it gets difficult to hold onto the expectation of 70F. The way it feels to set your "setpoint" -- or "expectations" or "intentions", if you want to think of it that way -- is by aiming our attention at the experience of that thing being true. This makes any deviations become salient, feel "not right", and motivates the response to correct them.

If we want to hit the target, we aim to experience the sights lining up and *staying* lined up through a smooth trigger pull and follow through.

If we want to grow up to be a doctor or lawyer and our report card comes back with bad grades, it conflicts with the fantasy we've been living in, and feels like "This can't be!".

What if we divorced this attention from *any* framing? Forget "intend", "expect", "imagine". What if we *just look* at a thing?

I tried an interesting experiment on this years ago, when I was trying to sort out exactly what it takes for a "hypnotic suggestion" to take. The hypothesis I was testing is that all it takes is to just put attention on an idea, and *not take it off*. So how can we do that, at its simplest and most detached from justifying contexts like "I wanted to do that"/"That's just what was gonna happen"/"I was hypnotized to"/etc?

I had an idea.

What if I just *told* people a thing -- for example, "when I scratch my nose, you will find it to be the funniest thing in the world" -- and then made up a weird excuse to get their attention off the fact that I just put their attention on something stupid? Maybe I could leverage "hypnotic amnesia" to do this? After all, if this hypothesis is right... I should be able to achieve *that too*, just by saying it and getting awareness off it.

So now the suggestion stack is "when I scratch my nose, you will find it to be the funniest thing in the world, and you will find it *even funnier*, because you will have no idea why you're laughing because you won't remember any of this the moment I snap my fingers".

Okay, so how do we get their attention off *that*, for long enough for the suggestion to take? How about just distracting them? Well, that's a start, but they're likely to *preemptively* keep their attention from focusing on that to the exclusion of "this isn't *real*".

So what I did is give them a reason to not think it was necessary to.

"I'm going to say something really absurd and not even true so I don't expect you to believe me, just hear me out and tell me *if you understand* -- okay? Okay. In a moment, when I scratch my nose, you will find it to be the funniest thing in the world, and you will find it *even funnier*, because you will have no idea why you're laughing because you won't remember any of this the moment I snap my fingers. Of course this isn't actually going to happen and I'm just saying dumb things, but *you do understand, don't you*. That's right, and <snap> now what were you saying earlier -- you remember what we talked about when you first got here, with the whatever and such?"

Pulling off the distraction wasn't exactly trivial, but when it worked... it worked. I'd scratch my nose, they'd laugh -- not like it's the *funniest thing in the world*, giving another hint about what it takes to enact "hypnotic suggestions" -- but they did laugh, and they did not know why they were laughing until reminded.²

Was it their "expectation" that they'd have amnesia? Or "intention"? Not really, in that there was nothing deliberate about it, and no evidence that it would be true or meta-beliefs in support. There was no bus, no tiger, they were *just running*. They weren't even "imagining" it, really, just holding that idea in mind, *for a moment* -- they thought. [We really do believe everything we're told](#), unless we actively move to disbelieve.

When you're doing your best to hold the temperature at 75.00F, notice what you *have to* look at. If you are to do a maximally good job at regulating the temperature, you have to be looking at that thermometer read out, or else you won't notice when reality shifts away from your intention. You have to be *focused*, so that irrelevant nonsense doesn't distract you from noticing the instant the temperature drifts to 74.99F. You are constantly expecting to see 75.00F in the proximate future, and acting so as to maintain this expectation. This is where your attention lies, and is the experience to which errors are measured.

At least to me, this is the most confusing part of the sequence. Because it sure seems like "But can't you just attend to things *as they are* without also trying to change them???". Like, can't you just watch the thermometer readout intently because you want to *know* the temperature, without also trying to *control* the temperature? Of course.

I'm not generally a fan of the meditator "Focus on your breath" thing, but it does make a perfect example here because experienced meditators *can* observe their breath without controlling, and at the same time it isn't exactly trivial.

When we ask questions like "What will our breath do?" (or "Will my attempt to not swell 'work'?", or "Will my science experiment find a correlation?"), in order to answer them on

purpose we have to intend to resolve our uncertainty. And so what happens, naturally, is that we do things that minimize this uncertainty. We come up with potential answers, like "This is what my breath will do", or "It probably won't work, expecting that would make me crazy", or "I think I'll find a correlation, that's why I'm testing it" -- and by shifting our attention to this expectation, we [inadvertently set our intention](#). The intention of resolving uncertainty flows downhill both to "finding solutions that work" and to *making* solutions work.

So how do you "Just observe" your breath, or "Watch the thermometer readout, without trying to control it"?

By practicing the skill of holding expectations lightly. If your attention sticks on "It will say 75.00F", and it drifts down to 74.98, 74.95, this builds tension, and this tension motivates actions to resolve it. If you have motive to hold that expectation, you may flinch and say "NO! It WILL be 75F!" and do whatever you can to make reality not be what it is.

If you don't have such motivation, then you can shift. "Okay, it's down to 74.98F. It'll probably continue down to 74.95F soon enough". This is easy enough when you don't have anything invested in the outcome³, but when you have reason to prefer one outcome to another *and* it's not in your best interest to hamfistedly try to control it, it can take some skill to allow your attention to track the data freely even as it provides that "[information that we care about](#)". The difficulty of observing one's own breathing without altering it shows how even subtle desires below the level of conscious endorsement can derail things pretty quickly, and often requires skillful direction of meta attention in order to keep our attention moving without friction or undue influence.

So yes, paying attention without controlling is possible -- *if* you are free of any motivation whatsoever to change the outcome or to even know the outcome. Often, this requires you to be attentive to any tension between your direction of attention and the data, and actively redirect your attention to maintain this lack of tension.

In other words, when you are expecting the thermometer readout to display 75.00F, that *inherently* functions as your brain's setpoint as the experience to create. In order to "just observe" you have to adjust your aim before you ever miss your target, generally by maintaining an awareness of the degree of error of your object level expectations, and aiming to drive this to zero. Sometimes this comes naturally, other times it very much does not.

Knowing this alone won't get you very far. There's a reason the "pure attention" example is a toy example and not something that was useful and truth tracking in a real situation. I had to pull some trickery to get that nose scratching experiment to work, and my attempts to distract didn't always work. People don't like reality violating their expectations and intentions, so they choose their attention... well, not *sufficiently* carefully all the time, but there's usually some amount of

thought that goes into it -- whether because they're intent on bringing about certain experiences or else preparing for them. That's why you generally can't just say "Think positively!" and expect that to work.⁴

But at the end of the day, attention is the final gear that drives intentional change, as it is the experience you attend to which serves as the setpoint of your brain⁵⁶. Everything else is either downstream of that, or it isn't intentional.

The rest of the question, is where is the right place to direct attention, and how do we get it there?

Notes

1. Usually, sneaky bastards that they are, hypnotists will pull some sleight of mouth to trick people into losing track of the distinction between imagination and reality. Sometimes they're upfront about it, and just say "Imagine that you're not imagining". Other times, they'll just lead with "Imagine..." and then slowly (or perhaps not so slowly) walk towards treating it as *real* while you're paying too close attention to what they're saying at the moment to notice that it contradicts what they were saying a moment ago about it being just pretend -- and then abruptly reminding you that you're just imagining when they want you to remember. ↩
2. This is something you could try for yourself, and might be able to pull off. I had a friend with relatively little experience playing hypnotist try it, and he got it to work on his second attempt (distraction failed on his first).
It probably does take *some* familiarity with the basic skills though. ↩
3. And foolish enough when you *do* -- and also have the power to change it ↩
4. I mean this in a technical/empirical sense. Try to expect it to work, and see if you can manage to be surprised when it doesn't. It's generally not easy to do. ↩
5. This is in line with Karl Friston's work on the brain as a free energy minimizer. ↩
6. As an exercise for the reader, what does this say about target fixation? ↩

Recognizing Optimality

Short feedback loops are great when the feedback is clear, but without a guiding framework it can be hard to know what to make of the feedback -- especially in the short term. If things get uncomfortable, is that a bad thing or is that a [counter-steering](#)? If I'm getting results quickly, am I [Goodharting](#) to rapid yet [unstable results](#)? If I'm not getting results quickly is it because the problem is hard or just that I'm ineffective? Am I even wise enough to know what a good outcome looks like in this situation?

A physicist can take the mass of your car, power capability of the engine, and the height differential between the start and finish, and give you a lower bound for how quickly you can race up a mountain. This is useful because if the lower bound is 99% of your actual time, then you know that you have little to gain in "learning how to drive" and need to focus on improving the power to weight ratio of your car if you want better results. If the lower bound is 1%, then don't bother upgrading the car because the answer is *all* in learning how to drive.

Even without the inputs to the equation the physicist can also judge your efficiency by watching your driving, and give criteria to optimize. If the accelerator isn't pinned to the floor, first do that. If you're revving past the power peak, upshift. If you're under it, then downshift. Optimality, in this case, looks like keeping the engine as close as possible to the power peak at every moment when the tires aren't at their limit of traction. If your engine is rarely near its power peak, you're not going to approach this lower bound. If you're up against the constraints, you're doing all you can. What might this look like when applied to "psychological" problems? What are the constraints?

To explore this, let's look at the maximally simple example of learning how to shoot a handgun effectively.

When learning to shoot a handgun there's various knowledge and skills that required. Once you get past the basics of things like "how to hold the gun" and "how to align the sights", there are still skills to be honed in order to do it well. You have to practice by feel so that you can keep the gun aligned and find your sights quickly. You have to practice to get a feel for what kind of sight picture is acceptable. You have to practice to get the feel of how to smoothly press the trigger so as to not pull the gun off target. And yet, even complete beginners can do a passable job at these things right at the start. The thing that produces really atrocious shooting is a failure of a different kind.

Shooting a handgun can be unpleasant. They are loud. Even with ear protection, they are still really loud. And startling. They kick back at you, and it can be uncomfortable. The hard part of shooting a handgun decently is in being okay with this and facing it without flinching. When you begin to press the trigger, your brain knows that this BANG is about to happen, and so if it interprets this "BANG" as a bad thing, it will change direction on you and reorient your focus towards damage mitigation. You push forward and down on the gun in order to brace for the recoil, and in doing so, you bring the gun far out of alignment. If the flinch is especially bad, you might even wince and close your eyes too, attempting to hide from the stimulus. This shift of focus from "smooth steady trigger pull while keeping the sights aligned and on target" to "Ahh!!! Protect myself!" at the last moment is the reason your shots fail to hit the target altogether -- at the moment it counts, you stop trying to shoot straight and start trying to do things that preempt that. It's a failure to even try.

So how do we fix a flinch? What is going on here, and how quickly can we expect to resolve this psychological issue?

Conventional wisdom is "not quickly". For example, [here's Navy SEAL Chris Sajnog on the topic](#). He is pushing back against people saying that it can't be done at all, and saying that it's actually easy and should only take about a month. In the comments, two people report trying his thirty day challenge. One spent 5 weeks and was successful¹, and the other was not successful². There was one more commenter describing their experience getting over a flinch (before watching the video), which took him 9 months of consistent effort³. Chris's suggestion is quite "open loop" both in that he tells you to not test for live fire results for the full thirty days, and that while he tells you to "tell your body you're a great shooter" he doesn't mention listening for a response.

Let's open the black box, and see if we can't do better.

A fairly common way to address flinching in new shooters, and the way I was taught as a kid, is with use of "dummy" rounds randomly interspersed so that the shooter does not know whether or not the gun will go off.

It can sometimes be hard to tell if you're flinching when the gun is going off every time. If the gun doesn't go off when you expect it to, there's nothing to mask your flinch. The only thing to look at is your flinch and you feel kinda silly for flinching for no reason. Busted. So if one out of every 5-8 rounds is a dummy round -- and you don't know which -- then when you get to it you will still mostly expect the gun to go off, and if you're flinching when you know the round is live, you're almost certainly flinching here. This gives you a way to catch yourself failing, so you get to notice and have that "Ugh, I messed up" corrective nudge. It also keeps you on your

toes *knowing* that you're going to get caught out if you keep flinching, so that not-messing-up stays on your mind.

If the gun goes off when you *don't* expect it to, you won't flinch because you won't have known to. If only one in every half dozen rounds is live, you'll still know it *might* go off, but it's much easier to refrain from flinching when you know it probably won't go off anyway, and the last three times you flinched all that happened was that you felt "Doh! I flinched for no reason again". By mixing in dummy rounds, you keep yourself in a state of uncertainty and unable to miss the results of your flinching, so it is a much more efficient method of learning to overcome one's flinch than simply "shoot a bunch" and hope that you end up getting over it by intentionally putting yourself in a position where you might accidentally learn. When you mix in dummy rounds, you get a surprise every few rounds at most, and [that surprise is where you learn something](#). Maybe it's "I can fire a live round without flinching", or maybe it's "I'm flinching even on duds", but you're learning *something*. It keeps you from the repetitive monotony of failing over and over and over without surprise, without corrective feedback, and without change.

What if we were to try to maximize this learning rate? If only one in ten rounds is a surprise, you're wasting most of your time. The first obvious thing to do maximize the expected information per round ("entropy") by using an even mixture of live to dummy rounds so that every trigger pull brings a surprise -- but even this is suboptimal. You aim, squeeze the trigger, and when the hammer drops you find out whether the round was live or not and whether you flinched or not. This means only part of the time are you updating on what you might be doing wrong, and the rest of the time you're flying blind.

Consider this instead:

Instead of pulling the trigger only to find out later whether you flinched, pay attention *now*. Do everything slowly. Slowly enough that you notice the impulse to flinch *before* you succumb to it. Only increase pressure on the trigger when you *know* that you aren't about to flinch. You won't know when the gun will go off, but the point isn't "Take so long that you won't know when to flinch". As you increase trigger pressure, you will go from very certain that that gun is not about to go off to less sure to heightened anticipation. You don't flinch when you know it isn't going to go off, you flinch when you think it will. Or might. By increasing trigger pressure only when you *know* you aren't flinching, you quickly get to that point of heightened anticipation and you get to spend the vast majority of your time sitting there, knowing that the sear is near releasing, knowing that the BANG is near, and maintaining control of your focus on holding the sights on target, and only increasing trigger pressure when you can do so without losing sense of priorities.

When you're on that edge, the gun is about to go off with an unknown-but-small increase in trigger pressure, and in that state of heightened anticipation, you have time to reflect. And instead of "Yes, actually, this is legitimately scary and deserves to be", you tend to realize "Actually, it's not that bad. I've been here before". The nerves fade, you calm down a little, and as you do you increase trigger pressure to bring yourself right back to that same level of nerves while being closer to the point of firing.

What happens as you iterate this process is that you get to continually sit in the conflict between "It feels scary to shoot a gun" and "I'm okay with recoil. I want to hit the target", and realign your focus as necessary, as your brain realizes which of these two is more important and more sensible to you. Eventually, you get to the point where the sear releases, the bullet flies out the end, and you repeat the process. As you spend more and more time doing this, you accumulate more time realizing that the recoil and report ain't that big a deal, and the nerves become less and less prominent. Each time it goes a bit quicker, and before long it's not necessary to slow down at all in order to maintain focus where you intend to keep it.

This is the process of reprogramming your own mind at optimal speed and precision. It involves relentlessly directing your attention towards [the experience of realizing your goals](#), and the moment this becomes difficult to achieve, stopping to notice what your revealed uncertainty is and what you want to do about it. Aside from a quick trigger reset, the entire time is spent in this "figuring out what the right move is" mode, and reconditioning your response as your best bet becomes apparent. Known-failure is never accepted, and you don't flinch even on your first shot. This isn't just "theoretical". It *actually works*, both on those who have been struggling with a flinch and those who have never fired a gun before in their lives.

Efficiently climbing hills means choosing a gear that keeps the engine near the power peak and spending the whole time full throttle, with a few moments here and there spent shifting. Efficiently changing minds means choosing a rate and type of approach that keeps focus where it needs to be so that the both sides of the relevant question are at the forefront, and spending the whole time there weighing the options -- with a few moments here and there spent resetting the trigger and deciding on new targets to address.

In practice, this often happens so rapidly that we can miss it. In the cases where I've explicitly taught someone how to not flinch while shooting firearms, there was no prolonged "I'm overcoming a flinch" phase, just a smooth steady trigger pull and no flinch -- from shot one. Shooting a gun just isn't that scary, and so it generally doesn't take but a few moments of properly aimed attention to address. Anything lasting much longer than that is the result of failing to hold focus on the unresolved question, rather than the result of having a big hill to climb.

Not all conflicts are so simple, but the metric for optimality is the same. Is near 100% of the time spent with attention on that relevant uncertainty that separates one side of the [disagreement](#) from the other, with only a moment here and there when shifting between sub-disagreements? Or are words and actions thrown around without expectation of anything *surprising* happening, that someone can learn from?

Notes

1. "Tried it for five weeks. Every day. Maybe 10 maybe 15 times. Found more than that lessened my concentration. Went to the range yesterday. 45 acp 9 out of 10 in the black (3" @ 30'). 9mm 10 for 10. Will continue dry firing every day. Bragging absolutely not more like pleasantly surprised. Thank You Chris" ↩
2. "Have a bad flinch. I tried this method and it didn't work for me. I'm not complaining. It might work for some people. This is just feedback." ↩
3. "Excellent suggesrions Chris! I worked hard about 14 years ago to overcome the flinch. I did LOTS of dry fire practice with some .22 rimfire practice sprinkled in. It took me something like 9 months of consistenr effort to work through it." ↩

Part: Attention

Attention is necessary and sufficient for all change, and figuring out where to place attention is the big question.

The Spectrum of Attention: From Empathy to Hypnosis

Whenever there's a "[disagreement](#)" -- again, including everything from the obvious "Who should we vote for?" to the physiological "Should we swell this sprained ankle?" to the 'unnegotiable' "Should you love me?" -- there's a question over which side deserves attention, and to what degree. Part of this is "Who should be talking, and who should be listening?", but this is making a binary out of a deep and rich spectrum.

The thing that makes hypnosis so bizarre and seemingly powerful is its ability to *keep* attention, in the face of ordinarily compelling objections. I'll give you an example.

Right now, you're reading what I've written. You've tentatively agreed to follow my direction of attention, for now. The moment I say something *wrong* -- or at least, something that *you think* is wrong, using your worldview to guide you, I lose that. If I tell you that when I say the name of a color, you will no longer be able to remember your name, you don't say "Wow... that's amazing, I wasn't expecting to be unable to remember my name!". You say "No" -- or maybe "We'll see about that", with an expectation of experiencing something else. Rather than following my lead, and attending to the possible world I describe, you reject my bid for attention and attend to a possible world that is mutually exclusive with the picture I'm painting. Which you should, because that'd be a dumb thing to say.

When "in hypnosis" though, people don't. Setting aside the "why?" for a moment, people who are hypnotized attend to the predictions given by the hypnotist. If the hypnotist predicts that when he says the name of a color you will be unable to remember your name, that is what you attend to, and that is what you expect, so that is what you experience -- because that is the experience you create. When you follow along despite being told strange and counter-worldview things, and decline to point your attention in an objecting direction, you experience strange and counter-worldview things.

Richard Feynman's experience with hypnosis demonstrates this well. Curious what the experience of being hypnotized was like, Feynman followed along:

He started to work on me and soon I got into a position where he said, "You can't open your eyes."

I said to myself, "I bet I could open my eyes, but I don't want to disturb the situation: Let's see how much further it goes." It was an interesting situation: You're only slightly fogged

out, and although you've lost a little bit, you're pretty sure you could open your eyes. But of course, you're not opening your eyes, so in a sense you can't do it.

He went through a lot of stuff and decided that I was pretty good.

When the real demonstration came he had us walk on stage, and he hypnotized us in front of the whole Princeton Graduate College. This time the effect was stronger; I guess I had learned how to become hypnotized. The hypnotist made various demonstrations, having me do things that I couldn't normally do, and at the end he said that after I came out of hypnosis, instead of returning to my seat directly, which was the natural way to go, I would walk all the way around the room and go to my seat from the back.

All through the demonstration I was vaguely aware of what was going on, and cooperating with the things the hypnotist said, but this time I decided, "Damn it, enough is enough! I'm gonna go straight to my seat."

When it was time to get up and go off the stage, I started to walk straight to my seat. But then an annoying feeling came over me: I felt so uncomfortable that I couldn't continue. I walked all the way around the hall.

This all seems like poor design from a blind idiot god again. Why design a mind so as to be so susceptible to nonsense?

Because that's how you get susceptible to learning that what you *think* is nonsense, isn't.

Here's the rest of Feynman's experience with hypnosis:

I was hypnotized in another situation some time later by a woman. While I was hypnotized she said, "I'm going to light a match, blow it out, and immediately touch the back of your hand with it. You will feel no pain."

I thought, "Baloney!" She took a match, lit it, blew it out, and touched it to the back of my hand. It felt slightly warm. My eyes were closed through out all of this, but I was thinking, "That's easy. She lit one match, but touched a different match to my hand. There's nothin' to that; it's a fake!"

When I came out of the hypnosis and looked at the back of my hand, I got the biggest surprise: There was a burn on the back of my hand. Soon a blister grew, and it never hurt at all, even when it broke.

On the one hand, "You can't remember your name!" is stupid, and false. "You will walk the long way around" wasn't in line with what Feynman *wanted* to be true, and it didn't have to be.

On the other hand, "You don't have to have the normal unpleasant experience of pain. You can simply be comfortable, if you want" is *true*. And important. And not something he was primed to notice without hypnosis.

The idea that you can simply *not be bothered* when a match is put out on your skin is *unthinkable* a lot of the time, despite being totally and completely true -- and if you can't think it, you can't act on it even when it would make sense to do so. Getting people to point their attention in that direction takes a more compelling bid for attention than people are used to making.

In Terminator 2, they depict the terminator as [not responding to a cigar burning his chest](#) as a way to demonstrate his inhuman nature. The difference here isn't that the terminator isn't capable of sensing damage. The difference is that the damage isn't particularly interesting to him, and does not control his attention so he's free to focus on other things. The actual sensations a human would feel in this situation is a barely scaled up version of what Feynman went through no problem, and this reaction is readily achievable if that's what makes sense to do. The big difference is that not being a terminator, if you were to try such a bold move in a biker bar, the rest of the scene might unfold differently *and you know that*. If you were to ever consider something so foolish, the rest of your brain would be screaming at you, and it'd be hard to ignore because the warning is clearly legit.¹

But if the damage has already happened, and you're not in a biker bar being threatened? When my thirteen year old cousin picked up that hot fire poker, that burn was worse than the terminator's. And yet, as soon as he thought about what the problem was, he wasn't any more bothered than the terminator. So maybe it's not our designer being less than intelligent, but us being less than intelligent.

This raises the question:

In what contexts *does* it make sense to attend to the things we might ordinarily reject out of hand?

Feynman gives a good hint:

I bet I could open my eyes, but I don't want to disturb the situation: Let's see how much further it goes.

It makes sense to focus on the seemingly absurd when the potential to learn things that are true and useful outweighs the costs of considering (and risking acting on) things that seem to be false.

There are multiple factors that can drive the decision here.

- Is it likely going somewhere interesting?
- Is it easy to throw up objections without interfering with the process of communication?
- Is your objection actually important to address right now or not?
- Am I at risk of harming myself by acting on false beliefs if I don't object?

And perhaps most importantly,

- Are you likely to be *wrong* about what you're objecting to?

Feynman's answers were "It would be interesting", "Objections would disrupt the process", "Postponing objections was fine", "Any false beliefs would be temporary and the harm acceptable", and "Maybe! It's almost a weird philosophical question I guess." So he let his attention follow the hypnotists bids and achieved the results he expected, to his surprise, as he thought might happen.²

With the swelling injury thing, my friend's answers were "I might learn something surprising", "Sure, but what can I say that he hasn't thought of", "Just try it first", "Worst case nothing happens", and "Yes, it's likely that my skepticism turns out to be wrong". So she had her skepticism, and didn't let it get in her way of listening and finding out.³

All of these questions are important, and there is much to say here. For example, there is value of listening to ideas *as ideas* rather than "as reality" so that they can be safely quarantined and allow you to consider more ideas for potential truth. Feynman's willingness to experience "being wrong" allowed him to test more ideas and find out first hand when some of them weren't. There is also great value in knowing what it takes for a fair test, and when it's important to hold off on stress testing until the full arc has been completed, so that you don't find that the ideas can't support their weight in your perspective when it's just that your perspective is missing a few critical pieces.

For sake of brevity though, I'm going to focus on the last one. Maybe we're wrong.

Us atheists have a terrible tendency to equate "faith" with "believing things that there is no evidence for", which is understandable, but ultimately an error into arrogance. The value in faith is that it allows you to believe *true* things, which there *is* evidence for, but which you cannot yet comprehend for yourself. "God works in mysterious ways" isn't just a cope, it's also a necessary fact about any wisdom generating process operating on a level above yours; for if its ways weren't mysterious, you'd be on that level too.

Consider children, for example. When children are young enough, they lack the ability to comprehend the danger of playing in the street, and simply explaining "you could get hit by a car" does not convey the needed understanding to make the relevant decisions wisely. The first time I tried explaining the danger to my daughter, she excitedly told me "And then we'd put a bandaid on the booboo!" -- she just had no experience with injury so bad that it justifies this degree of fear and respect. While helping her build her own understanding of things is an important long term goal, I can only do that if she's alive to build understandings. The more time sensitive thing is that she understands that streets *are* dangerous, not *why* they are dangerous.

The proper response, when in the position of a child, isn't "I can't see any reason to believe playing in the street could be *that* bad, and if we were to get a booboo we'd just put a bandaid on it". It's mama and dada seem *very serious* about this, I don't fully understand why, and despite the fact that I don't know *why* I do know *that* it is very important to not play in the street".

The proper response from the position of the adult isn't "Yes, you're totally capable of making the right decisions here, so long as you understand these object level facts". It's "I know that it might not fully make sense yet, and still, it is important that you stay out of the street".

Facilitating good navigation here necessarily requires that the child take the "god like" parents' view *on earned faith*, so that the child can believe in mysterious-*to-them* truths and use them to avoid tripping over things they don't see. The same processes which, when used inappropriately, leads to "Controlled to walk the long way by a [playfully] tyrannical hypnotist" and "Accepting false beliefs without evidence because someone speaking on behalf of a non-existent entity told them to", when used appropriately are just following a greater wisdom than one's own, when there's plenty of evidence to support this action.

My own father was very effective at this when I was still young enough to lack the proper perspective. The man let us play with firecrackers, when hysterical parents would freak out over the "danger" of maybe getting a blood blister if you're some combination of dumb and unlucky. To the extent I may have been blind, I had plenty of evidence that it was worth taking his perspective here on faith; if there wasn't reason to believe cars posed a very serious danger, he wouldn't be acting like they do. As a result, I listened, and behaved in accordance with his perspective that I couldn't yet fully support on my own.⁴ While the full extent of this might not

make sense until the "security" section, another word for this "looking to understand others' perspectives as true" is "empathy" -- it's taking on another's experience as one's own. I was noticing "This seems really important to him", and feeling the importance of that bit of his experience.

This dynamic of trusting and following guidance beyond our current understanding often comes up in hypnotism, where people will ask hypnotists "Do you hypnotize your kids?" -- or even "Do you hypnotize your wife?". Answers vary from "Oh no, I would never" to "Of course ;)", but the best one is that it's a [confused question](#), and that what my dad did with me is hypnosis in the most meaningful sense. I'll explain by answering "Do you hypnotize your wife?" with an example.

One morning my wife woke me up asking for help in a concerned tone, with a stream of blood pouring out of her hand. She had dropped a glass and cut one of the larger veins in the webbing between her thumb and forefinger. I bunched up some paper towel and put pressure on the cut, raising her hand to decrease the pressure. And then, since we were stuck waiting for the bleeding to stop anyway, decided to try something I didn't expect to work.

I had heard of hypnotists stopping bleeding through vasoconstriction, so I thought about what it would be like to experience that working, and started explaining what I wanted her to do.

You know that feeling you get when it's cold out and your brain shuts down circulation to your hand to preserve warmth, and your hand just gets cold and numb? You know that feeling, yeah? You know what this feels like? Good. Bring that feeling back. Go there now.

From the outside, as was my position when instructing her, it probably feels like this shouldn't work. The ~~magic happens~~ mundaneness becomes apparent when you step down from the meta level intuiting whether this "would work", and intuit *what would happen*, directly, at the object level.

You do know that feeling you get when it's cold out and your brain shuts down circulation to your hand to preserve warmth, right? Like, you've had experience being out in the cold, and know that this is a real thing, with a sense of what this experience is like? And that it results in your hands being both *cold* and a degree of *numb*?

Does it make sense that when "thinking about what that's like", you begin to revisit that experience? Much like if you're asked what it would be like if your phone were to ring and it were to be an unexpected call from your crush, you could actually feel which emotions would come up in that situation, because you're simulating that situation and feeling them now? And that the limit of how fully you experience these emotions is likely to be your motivation to feel

and get your hopes up for something that isn't even real and contextually important? And that if you *were* to see it as contextually important to commit to -- for example, if you have a real world date coming up, and the anticipated scenario *will actually happen*, you tend to feel that anticipatory excitement/anxiety/etc almost as if you were there, and not just in a "Yeah, I guess I can kinda tell that I'm going to be excited" sort of way?

Does it make sense that when you revisit an experience, your body responds in kind, much like it does when you revisit the experience of biting into a lemon, and when you get the very real butterflies in your stomach anticipating the date that doesn't start for another thirty minutes?

When you're asked to revisit the experience where your hand gets *cold* and *numb*, because it's cold out and your brain is shutting down blood flow to that area, does it make sense that you could revisit that experience, somewhat, when you have sufficient motivation, and that blood pouring out of your hand might be a little motivating? Does it make sense that in doing so, your brain might start doing the same things, leading to some amount of the same physiological responses?

We held pressure on the cut for a minute, maybe two tops, and when we pulled back the paper towel, the wound was dry. Not only had it stopped bleeding, there was no blood visible -- like when you cut into a chicken breast from the store. Not solidified into a big clot, just not there. I was shocked, and had never seen anything like it.⁵

When the doctor was stitching her hand back together, he noticed how cold her hand was and expressed his concern that she had circulation issues that needed to be addressed. I told him to feel her other hand, which was fully warm, and explained that I suggested she constrict her blood vessels to stop the bleeding. He went back to stitching without comment.

In one sense, that's a clear "Yes, I hypnotize my wife sometimes". Because I achieved a "hypnotic phenomenon", the pattern of pacing and emphasis is something I picked up doing hypnosis, and no hypnotist worth their salt thinks that formal "inductions" are needed. In a more important sense, literally all that happened is that I explained to her what to do, and given the circumstance and our history she was inclined to listen and intend to do what I instructed.⁶ It may be an unusual target suggestion, and I may have learned to extend into the "hypnotic" side of the attention spectrum from practicing more formal "hypnosis", but the fundamental dynamics at play are the exact same as when my dad explained not to play in the street. So called "hypnotists" do have their conspicuous grab bag of "tricks and techniques", but the thing that matters -- the thing that determines whether their suggestions take or not -- is whether they succeed in directing attention towards the desired suggestion. This is what hypnosis *is*, in the most fundamental sense -- a willingness to put aside the normal "But that won't work!"/"But I can't do that", *shut up, and intend to do the impossible* merely worldview incompatible. It is

focused attention on one idea to the exclusion of others, no more, no less. Even James Braid, who popularized the term "hypnosis" which has stuck to this day, recognized not long later that "monoideism" is a much more fitting label.

Whether we respond to an idea -- for example "don't play in the street", or "constrict your blood vessels", or anything else -- is a matter of whether we allow it to hold our attention or whether we dismiss it and look elsewhere. So the question is of what is worth attending to.

The more we know what is *true*, and important to see, the more we are called to look at what is true, point at what we are seeing, and expect others to accept our invitation to look with us -- even if it might seem "crazy".

The more we need guidance, and have available someone who can see the things we cannot, the more we are called to *listen*, and follow the gaze of those who see more -- even if we're not yet sure how to make sense of it.

In my physics program, things worked quite wonderfully, and the back and forth flowed fluidly. Most of the teaching was done by the professors, but occasionally the prof would say or write things which didn't quite make sense. In those cases, in the service of broader learning, the students would then speak, and point to what seems to be an error. Locally, the students were calling for the prof to look with them, and the prof would accept this guidance -- most of the time resulting in learning of their own error and acknowledging it, but sometimes learning where the student needed something explained, and explaining the missing piece so that the class could understand his lesson.

Sometimes though, the possibility of coming up with seemingly good objections is known, and not indicative of a problem in what's being taught. In those cases "Hear me out. I know this isn't going to make sense at first, and I promise it will all come together in the end" can be a lot more appropriate -- and lead to experiences like Feynman's where he's willing to hear things like "The earth is a giant spherical rock" without "Um ackshually"ing that the earth isn't quite spherical or just rock, and following along until he learns something true and so surprising that he had no way of fitting it into his existing conceptual framework.

This is the spectrum we have to play on. Sometimes, in the moment and near future, its important to stfu and listen because any objections we could come up with would be irrelevant or mistaken, and experience truth straight from the source. Sometimes we need to be vigilant to a degree, and point out when we have objections we *do* expect to be worth addressing -- and find out. Sometimes we're on more even ground and attention flows from one thing to another being directed more or less equally from both sides. And sometimes we're on the other half of the spectrum, having things to point to and convey, with various levels of sensitivity to objection being appropriate.

The challenge isn't in "How do we get people's attention so that they do what [we think] is right" but in "To what degree do we even think we even know what is right here?". In order to learn physics, a student must evaluate when it is worth voicing that something doesn't make sense -- and to teach the teacher must evaluate whether it's worth addressing the confusion on the spot, or whether to trust it will resolve as they get further into the lecture. In order to help my wife constrict her blood vessels, I had to evaluate whether it was worth instructing her to do a thing she might not be able to do -- and she had to evaluate whether it was worth pointing out that this is kinda a ridiculous ask. We constantly -- on a moment by moment basis -- face these questions: to what degree does it make sense for them to look where I'm pointing, and to what degree does it make sense for me to look where they're pointing?

In full blown hypnosis, I can say "You will not remember your name" and it just doesn't matter whether that's absurd or not because they are putting their attention where I specify without doubt or hesitation. Everything I say becomes their reality -- as hypnotists like to say in "[the super suggestion](#)". This is rarely appropriate, but approaching this can have its place -- especially for a moment at a time.

In what might be called a "lighter trance", the "hypnotee" -- or "student", or "learner" to be most general -- will go along with most things, but object when things get too far from fitting with their conception of the world. The sensitivity to objection has been turned up from epsilon to something finite.

As we continue to crank this knob, the "student" becomes the "teacher", and the teacher begins to attend to the student to increasing exclusion of what he was going to teach, until we're on the other side of the spectrum and the hypnotic attention flows in the other direction.

Notes

1. As an aside, this scene makes a good example of quickly escalating from attention to respect to security, and then achieving agreement on the fact that they were going to give him their clothes.

Terminator asks for clothes, holds his attention there bidding strongly. He is immediately laughed at, showing that his bids are rejected due to insufficient respect, and the biker counterbids that terminator attend to the cigar burn on his chest. Because the burn isn't threatening to him, his attention isn't pulled away from his initial focus, surprising the biker -- giving him *new information* and causing him to *learn something*. The bikers are then pushed into insecurity and fight back out of fear (they're afraid of losing what they'd lose to give this guy respect, e.g. clothes/etc). Then they decide that giving him respect is actually the safer thing, so they give him the respect and therefore the clothes. ↩

2. This is not a contradiction. He expected to walk the long way around, so he did. He didn't expect that this expectation would be fulfilled, so he was surprised when it was. He expected that some of his (meta)

expectations would be falsified, and just as he thought, they were. The objects of each expectation are on different meta levels. ↩

3. I should explicitly say that I do *not* predict that you will be able to pull it off if you "try" to not swell your injuries. Maybe you would, maybe not. Like the flinch example starts to show, it can be non-obvious what one is actually trying to do sometimes. I was able to tell her other true things about what she could do that she didn't realize that she could do, and have her *predictably* fail to try and fail to notice that she wasn't trying -- even after getting this one right ↩
4. It's worth noting that this often doesn't *feel* like "I don't get this at all, but okay". It felt like "Okay, makes sense" because the reasoning he gave did make sense -- as teacher's passwords and other insufficient understandings often do.

But would I have recognized and appreciated the gravity of the situation on my own? The idea that a thing "makes sense" usually functions more as a justification for keeping an embodied belief than it does to *create* the belief. The latter comes straight from the direct suggestion to stay out of the street. We try on the conclusion first, and then use reasoning to decide whether to reject it or justify it. ↩

5. Except, I guess, when I'd cut into chicken breast from the store. ↩
6. This pattern of "Hey, look what falls out of this framework naturally!" will turn up again and again. In cases where you know the right answers, and they're counterintuitive, and the normal objections don't matter because what's important is making space to see the world differently, then you get focused attention.

When you navigate this way, and end up in these situations, you can be mid conversation and notice "Wow, the person I'm talking to is *hypnotized*" -- and then recognize why it makes sense in context. ↩

Navigating Attention

In the [last post](#) I introduced the concept of the direction of attention existing on a spectrum. On one side of the spectrum, we set aside what would otherwise be salient objections and just *listen* -- aiming to experience the world as they explain it. We might call this "empathy". On the other, they set aside objections that would otherwise be salient *to them*, and aim to experience the world as we explain it. We might call this "hypnosis", though it's just the other wide of the same coin.

In this post, we'll go over how to actually navigate this spectrum, and what the challenges are. It seems weird to be able to say stuff like "Jumping off this cliff is safe" and have that *actually* quell fear, let alone "turns out you can just decide to not swell injuries" and have that lead to injuries not swelling. So why doesn't that *always* work? How do we ensure that when it should, it does? The default expectation is that "just saying" things like this would lead to people still being afraid *even if* they believe you. "Fears are subconscious", "I believe you but I don't know how to believe it", and the like.

A recurring theme of this sequence is that this is all simpler than it seems. [The way to hit the target when you have a flinch is to intend to hit the target](#). That is, attend to the experience of hitting the target, and act so as to correct things whenever evidence points to you getting off track. The flinch is a distraction, both figuratively and literally. And the way to get past this distraction is to look at it, and decide whether it's worth attending to.

Similarly, the way to navigate direction of attention is to attend to the question of where our attention belongs, and this dramatically simplifies things like "How to get people to listen to us, given that we're definitely totally right?". And similarly, this application of "just look at reality" calls for some explanation.

In this case, the distractions are the [semantic stop signs](#) like "I know, I know, but my brain doesn't get it". Just like you have an opportunity to notice that you can handle the recoil, and decide to aim at the target, you can notice that the semantic stop signs are serving to keep you disconnected from the truth, and decide to aim at the truth. "Okay, I feel anticipatory nerves about this recoil. And still, I will aim at this target". "Okay, my brain doesn't get it yet -- or whatever. That's why I will stop and *look*".

Once you do that -- on either side -- you get to aim at the thing that matters, and get results that may have felt out of reach. When you have clarity on what's worth looking at, and to how much exclusion of competing shiny objects, the mechanics of "hypnotic inductions" are kinda irrelevant. You still have to do the mechanics that work, but because it's all downhill from

established intent there's no need to micromanage your own behaviors. The behavior functioning as an "induction" -- that is, the behavior that conveys the "This is important, none of that other stuff you're looking at matters right now" cues -- can be as simple as "Shhhhh...". Or a look. Or silence, while you wait until they notice the gravity of the situation. Or whatever else might be contextually appropriate.

Let's look at how this work, and how it doesn't, starting from the idea that we're definitely totally right.¹

Let's pick a "payload suggestion" -- that is, the object level idea that all of our other communication will be in service of, and the thing that creates the desired ("correct") behavior when attended to.

What might that be?

Maybe "If your makeup washes off, *that's okay*".

Maybe it's something we want them to do, like "[You're going to] pass the salt", or "[You're going to] constrict the blood vessels in that hand".

Or maybe the person is in pain, and you want them to not suffer over it. So maybe "The pain is gone. It just doesn't hurt".

Great, so we know what we're trying to convince them of.

Is it true?

Is it okay if her makeup washes off? Will they do as you instruct? Is the pain gone?

It's easy to respond to these with cached answers. "Yes, dammit!", "They *should*!", "It will be if you listen! It's self fulfilling!". But these are all flinches, which disconnect you from your *actual prediction*. Close your eyes, and play the sim forward the way you would in your dreams. *What happens?* What is it that you observe yourself anticipating?

Often the answer here won't be what we want. She probably won't believe us enough to try. They might not do what we want. Our "pain cure" might not "work".

These are distractions as well. Another recurring theme of this sequence is that *if you're actually justified*, you can "just say things" and people will actually believe you². Not because of magic and maybe not "just", but the justification that allows you to *know* the truth of your words also allows you to give that look that comes off as credible, and to say "Shhh..." without it feeling disrespectful or like an attempt to manipulate.

So let's set these aside too, for now, and look at what happens when we *do* have faith that if we're correct and justified we will be believed. This is where it gets interesting, because it leads to the question that lies at the heart of rationality.

What if I'm wrong?

It's really easy to dismiss a fear as irrational, when you know the irrational person won't listen to you anyway. It's easy to tell the girl in the jacuzzi "You're fine!", when you know she's not going to listen to you anyway. The moment you get the power to overwrite beliefs with your word, when people are willing to take your words at face value and take risks on your word alone, things get *scary*. It's sobering.

We start to realize we don't want to be wrong, and this shapes how we think and speak. When you hold yourself to giving instructions and asserting truths only when you, yourself, expect them to be true and complied with, your unqualified statements become much rarer, and much more compelling.

Are you *sure* that cliff is safe for your friend to jump off? For them, in the way they'd do it? At this time, in these conditions? The cost to be afraid of isn't that they don't believe you, it's that they *do* -- and that you might be wrong.

Are you *sure* that the girl won't be judged for her looks, in a way that is more traumatizing than she can handle? *Is the pain gone?* Then why are you lying to someone you purport to care about? Even if you notice to shift to "You don't have to attend to this pain", are you *sure* the pain is safe to ignore? What if you're wrong? Are you ready and willing to have blood on your hands?

Like the previous question, often the answer here isn't what we want to think either. The fear is *probably* not needed. Or overblown, at least. Probably. Maybe. But that "What if" is a bitch.

Until you accept that responsibility and feel the weight of the decision as your own, you're not speaking conclusions -- only arguments -- hoping to sway the person into making that decision *themselves*, so that the responsibility lies on them should it be in error. We can never be 100% sure of anything, and so if we aim to lead someone we may end up leading them astray. Are we willing to face the risk? To say "I know I might be wrong, and misleading this person who has placed their trust in me, and *still*, I want that to be on me if I'm wrong"? To answer the question "What if I'm wrong?" with "Then these are the consequences" *without* any distancing/minimizing language, to separate us from the truth?

When we can answer that question honestly with "Yes. I am willing to accept the consequences of focusing on this", then we can expect people to believe us when we bid for attention on the

"payload" suggestion. Until then, our attention is split between "I think this, and yet I worry that" -- and our bid for attention is polluted by that worry. As if to say "Trust me! But not enough for it to be my fault, in case I'm wrong!". As a result, the bids we made will necessarily be for unfocused attention, calling for significant attention to possibilities that our advice is in error.

Having faced the question of "What if we're wrong?" and accepted the consequences, we've made up *our own* mind about how much attention should be focused and in which direction. We're no longer flinching, and are now free from internal resistance to bidding for focused attention. Again, the particular ways we do it aren't especially important. Maybe you explicitly ask for attention, and for them to put away distractions. Maybe the lack of tension in your mind shows in the lack of muscular tension on your face, conveying the message. Maybe the cadence in your voice shows the weight you give the statements you make. Maybe you don't know how you are communicating your intent, but when your intent is on conveying a message and you expect to be understood or believed, you *tend to find a way*.

When we find that we *aren't* willing to accept the consequences wholesale, unconditionally, we can't bid so strongly for attention because we know we don't want our bids to be accepted so unconditionally. So we qualify -- intentionally or otherwise. Instead of "I got you. You're safe", it's "I got the rope. Did you check your knot?", or maybe "I got you?". When that's not enough, that signals something else might be going on that we're missing. Do I not have the [right] rope? Is there a problem (or potential problem) other than the knot and the rope? Maybe not, but maybe.

And so this brings up the other important question for navigating the layer of attention: "What am I missing?"

This question does two things, depending on the answer.

Sometimes there's an identifiable answer. When there is, you get to notice that you don't understand their perspective well enough. Maybe you don't realize it's the concrete anchor that's shaky. Maybe you don't recognize what it's like to be an insecure young woman opening herself to being judged. Maybe you'd be singing a different tune if you did.

When the degree to which we can ask to be listened to isn't strong enough to compel listening, the other option for engaging is to listen. Sometimes it's easy, like "It's the concrete anchor, you idiot", but sometimes it's not. Sometimes it's fear, and "I dunno! I'm just scared!", and you have to "listen" in that you focus on the truth of "I'm scared", accept that as a meaningful statement, and explore *what it means*. How would *you* feel, in those shoes up there, saying these things. Experiencing the compelling danger as *meaningful* and *potentially real* -- and, if that doesn't just seem right, exploring where it breaks down. It's not about the rope. It's not about the knot. What

could it be pointing at? Sometimes it's hard because it's illegible, and it's on you to [bring it into focus](#) and understand.

Other times it's not easy because what you're missing isn't what you think it's supposed to be.

You: What am I missing?

Other person: I'm not gonna tell you.

~~You: I guess I can't listen then~~

You: Hm, the thing I need to listen to is a meta-level unwillingness to even approach the object level. I wonder what *that* is about.³

And sometimes the only answer is "...nothing".

If someone is trying to tell you how to live your life, and you don't like that, it's easy to mindlessly slip into an attitude of "You don't know me!". And it's easy to brush off "Yes I do!" because "Pshh, he don't even know how much he don't know me". When they stop and sincerely ask "What am I missing?", it's an opening to tell them. You get to tell them off, and justify your refusal to entertain their ideas, so fairly inviting. And when the answer is "...nothing. I guess you do get me, huh", then with what are you going to support your disinterest in listening?

Even if there might be something there, the willingness to hear what they're missing conveys something about how informed their input is likely to be. It's not a result of pushing our perspective away. Maybe they don't know *everything*, but maybe what they do know is worth listening to.

On the level of attention, we can bid strongly to the extent that we can answer "What if I'm wrong" with "I'm willing to eat that cost" -- though the degree to which this will translate to *them* being willing to eat that cost rather than breaking rapport depends on the finite resource of "respect", so bidding hard risks escalation.

Alternatively, we can gain attention by offering attention once the other party decides that we're informed enough to have a perspective worth listening to.

Until then, we get to progress towards a solution by offering attention, and learning what it is that -- in their perspective -- we're missing and is of importance. Because by that point, we're in agreement that we're likely missing something important.

Notes

1. Of course, none of us would actually *act* on such an assumption, since we're all perfectly genuine collaborative truth seekers here. But we can pretend, and remind ourselves of how this assumption breaks down.

And sometimes we actually are right, and fairly predictably, so it's not entirely crazy. ↩

2. With a pretty big caveat here relating to security. ↩
3. I can't tell you how many times I've responded to people telling me they don't want to get into things with "That's okay. I understand", only for them to *immediately* open up about the thing they were just closed to. They just needed to know that I wasn't going to push them, and listening to that for one moment and reflecting it back is sometimes all it takes. ↩

Part: Respect

Respect is what backs bids for attention.

The Role of Respect: Why we inevitably appeal to authority

There is a large class of true statements which are indistinguishable from nonsense, and this is a problem for rationality. There's also a large class of statements which are *obviously* true, yet which we don't know. This also poses a problem.

Because if we're not just wrong, but *so* wrong that we can't even recognize what rightness would look like -- or what our own beliefs are -- how can we recover from that?

"Pain is just information" is one of those statements, which ordinarily gets brushed off as "*not true*" -- or "Not true in any meaningful way", at least. Yet I was able to dissolve my kid cousin's suffering in [a single word](#) when others had failed, simply because I knew that pain is just information. You can watch that knowledge lead to a resolution of suffering in the [chronic pain transcript](#) as well, where the conversation ended with the guy with debilitating pain from nerve damage telling me "You helped more with a couple messages than doctors have in over a year".

"You can just decide to not swell" is another one that seems *obviously* false. "It doesn't matter if your makeup washes off", to someone experiencing anxiety over it, could go either way. They might take it as "Obviously false", or they might take it as "I know I know, but knowing that doesn't make my brain stop". What normally *doesn't* happen, is "Hm. You're right. I think I'll go swimming now".

Yet that girl did go swimming, and found that it didn't matter if her makeup washed off. And my friend did decide to not swell her injuries, and found that she was in fact capable of deciding such things.

So wtf?

If these things are true, why do they seem so obviously wrong? If we "don't know" these things, then why does it feel so damn obvious that we do?

How many other things are like this, and how can we navigate these things intentionally, so that we don't get stuck with bad outcomes as a result of these false beliefs?

In the posts on attention ([1,2](#)) I talked about the part of the solution where we expect people to set aside objections and *listen* when we know we're right -- and where we expect *ourselves* to set aside objections and listen when we know they might be. What those posts don't cover, is when and why that process of negotiating attention breaks down.

If we're to imagine we know "It doesn't matter if your makeup washes off", and try to negotiate for attention by "just pointing at what's true" when some stranger is worried about it, what happens?

Maybe your anticipations differ from mine, but I don't anticipate that telling her "It doesn't matter if your makeup washes off" would get a response of the sort "Hm. I disagree. I think it's important enough to worry about". I don't expect any amount of "I disagree"/"I'm still confident" to ever get to "Okay, I agree" because I don't anticipate any Aumann updating happening. Common knowledge of each others rationality is a prerequisite for Aumann agreement to work, and I would expect her to quite reasonably view me as irrational if I were to expect her to take my word on that.

In other words, the response I anticipate is of the sort "The fact that you assert this isn't even evidence, so you don't have a seat at the attention negotiation table".

In other other words, "You aren't an authority on this"

The necessity of appeals to authority

Appeal to authority is generally considered to be fallacious¹, because the veracity of a statement doesn't change depending on who is saying it. False statements don't become true when a smart person says them, and true statements don't become false when a dumb person says them. Therefore, the argument goes, we should look only at whether the statements are true and the arguments valid, and not at who it is that is saying them. Sure, veracity of claims might *correlate* with authority, but not perfectly. So just look at what's *true*.

And that works great, when we're competent to figure out what's true. That there's value in this I think goes without saying here. The part that's a little more subtle is that as a general rule we can't actually do this. We cannot, as a general rule, simply look at the object level arguments and determine object level truth to the degree that the origin of the statement is entirely or almost entirely screened off.

Part of this is that even in things where we are quite good at figuring out what is true, often other people are too and so if there's a disagreement with the right other people it tends to shake our confidence. How many things do you think you know where you can honestly say that if all the smartest people were to tell you that you're wrong, you could confidently tell them "No, you are" without even a niggling doubt of "Am I missing something here? What the heck!?". For how many things would it be *rational* to take this extreme confidence? Even if it's something as simple as "Can I move my arm", is it infinitely more likely that the whole world went crazy, or is it possible that you have anosognosia? Even when we're looking at the object level and "thinking

for ourselves", we're necessarily supporting this with a degree of social validation in that if the right people were to disagree in the right way -- even content free -- we'd have a different confidence level. In this sense, we're never *purely* thinking for ourselves -- and that's a good thing.

The other part of it is that evaluating arguments on the object level is *laborious to the extreme*. It's not as simple as "I'd just check what's true" because as a general rule, "checking what's true" isn't a one step process like that.

For example, say we're out at a park somewhere and some guy is getting visibly and increasingly agitated, and angry with the person he's talking to. You look at him and think "This guy is going to hit someone". I disagree, and we agree to wait for more evidence before concluding. The next thing the dude says is "I'm going to punch you in the face!", and you excitedly say "See! See! The evidence supports me!". Does it really, though? Is it true that relative to the entirety of what we know, his verbal threats make it *more* likely that he's about to punch someone?

It's certainly possible. People often get increasingly worked up before they get violent, and frequently follow through on threats. The fact that he progressed to the point of yelling "I'm gonna punch you!" could definitely be evidence of that. It's not the *only* story that could fit though. Maybe I know the guy, and rather than some schizophrenic homeless guy struggling to hold it together, he's a fairly successful manager somewhere who has a reputation for a quick temper kept in check by an uncanny ability to figure out just how far he can go and get away with it. In that case, I might hear the same sensory data as you and think "Oh good, he only thinks he can get away with yelling here. If he thought he could get away with punching he'd have just done it". Which one is right depends on which model better explains why he's yelling, and we can't "just look at the evidence" until we know *what to do with* the evidence.

When our disagreement is at the surface, then yes, we can "just look at the data". If it's only one layer deep, then yes, you can say "I think he will yell on his way to punching", I can say "I think he will yell because he knows he can't punch", and having shared the differences in our models, we can be united in looking for evidence of whether he's more like a socially calibrated bully or a man losing his composure. If it's two or three layers deep, we can conceivably sit down and hash it all out -- if we have the time for that, *and* sufficient epistemic caution².

We usually don't have such clean and shallow differences in perspective though. You can have fairly deep political discussions where the same set of facts are interpreted and justified differently layers and layers deep, past where everyone loses track of their own justifications. You can't "just stick to the facts" with someone who disagrees with you politically, because in their mind all your sources have been debunked, while in your mind all of their debunkings have been debunked and vice versa. Look at the [lab leak debate](#), for example. Not only did that go

back and forth without resolving to agreement, [there were new layers that came up after the debate that neither person seemed aware of](#). And that [makes the whole debate fairly pointless](#). Or at least, not nearly as pointful as it could have been.

Even on things with few explicit layers of reasoning, often the reasoning chain is long and illegible. Whether you get a gut sense that the guy is bluffing or losing it depends on your entire history of social interaction -- including internal interactions with yourself, and including watching but not quite being able to put a finger on subtler and non-violent bluffs -- so when it comes to discussing why we predict loud threats *instead of* violence, we can't always say "Oh, I know that guy and he has a reputation". Sometimes we'd have to say "Okay, here's my whole dataset of all my experiences", which we obviously cannot do.

We're left with respect, or left with nothing

What we're stuck with, is estimating how much it means to us that the opaque box of another person's reasoning (or our own opaque box of reasoning, when our "gut" is telling us something we can't find rational reasons for) is giving this output, and using that to determine how far we're willing to entertain their conclusions. At low levels of respect, communication is *exhausting*, because every update must by necessity be infinitesimal, and building to anything surprising takes iteration after iteration.

What this feels like is when a crackpot tries to explain to you their new theory of everything, and cites their sources like Crackpot Weekly. It's hard not to laugh at everything they say, and if they were ever to be right about something as surprising as the cubic nature of time, and Crackpot Weekly actually pointed to valid arguments, actually, you might think you'd notice. You probably wouldn't.

On the other side of that coin, this is when you try explaining to someone why their political views are wrong and cite science -- only for them to laugh in your face, and ask "Don't you know that study has been debunked!?". And then when you look up the "debunking", it's in Crackpot Weekly and is utter trash. In order to get through to them with this level of respect, you have to do the leg work of finding *all* of their sources of support, how they justify them as valid, exactly where the debunkings go wrong *according to their own principles*, etc etc etc. You can be entirely right, with completely valid arguments, and they'd never notice. Because the disagreement runs deep enough that they won't notice that they're evaluating the validity of your arguments incorrectly. In some cases, if you're right you can just build the perpetual motion machine and start selling energy back to the grid. Often there's no clear cut proof in the pudding.

Look at Mormons, for example. South Park's caricature of them is as great people, happy, and quite successful -- e.g. the kid is a state champion wrestler. In my limited experience, South Park is eerily accurate; the Mormon family I knew were great people, happy, and while the kid wasn't literally a state champion wrestler... he was one brain fart away from becoming one, in double overtime, in one of the toughest states in the country. Pulling that kind of thing off often enough to be caricatured as such in a fairly anti-religious South Park sure looks like real world success to me. Yet I don't believe in magic underpants.

On the high respect side of the spectrum, these constraints are all relaxed. You can say whatever you want, knowing you won't be written off. You may be *disagreed with*. The person might not find your perspective persuasive, after trying it on. And you might be wrong. But you can get your fair shake, if you want. "Turns out you can just decide not to swell injuries" was met with skepticism, but an intent to actually test "I can decide to not swell this", not just "I can feel like I'm deciding and fail".

Schrödinger's Koan

"Koans" make a good illustration of what it's like to work with high degrees of respect. In western culture we like to laugh at the idea of some Supremely Wise monk sitting at the top of mountain, offering Great Wisdom like "The journey of a thousand miles begins with a single step", "Before enlightenment, chop wood, carry water. After enlightenment, chop wood, carry water", and "There is no self".

No shit, sounds pointless bro, and I'm pretty sure I exist -- respectively. So when Yoda of the mountain tops expects me to ponder the sound of one hand clapping, no thanks. You need two hands to clap, next question.

But that's kinda foolish.

Sure, it's easy to pose as "deeply wise" and just say shit without having great wisdom. Eckhart Tolle comes off that way to me. So far as I can tell he just suffered so much so pointlessly that he broke his ability to give a shit, and overshot in the other direction. So he spent a bunch of time sitting on a park bench doing nothing but giggling blissfully. That's not wisdom, that's poor regulation. That's him being slow to learn something everyone else got by default. Or maybe I'm wrong, and I'm missing his deep wisdom; I haven't read him carefully. But that's my point: I could be wrong here. The fact that superficial markers of deep wisdom can be faked doesn't mean that wisdom calling for that kind of respect *can't exist*. It means that we *can't know* whether it's deep wisdom. At least, not until we observe the meaning -- which can sometimes take years, if it happens at all.

Surely it was good that I listened to my parents about not playing in the street as a kid. I'm obviously smart enough to know that even long journeys start with a single step, but have I been smart enough to weight this fact *appropriately*? Would attending more to this fact lead me to just start stepping instead of getting overwhelmed by trying to take on the whole thing at once when it's not necessary? Maybe. Might there be something desirable about "enlightenment", even if it doesn't negate the need for chopping wood and carrying water? *Of course*. Might there be a purpose for focusing on seemingly absurd questions like the one about one hand clapping? I dunno about that one. Maybe if I pondered it long enough I'd find something. I decline to find out in this specific case.

Which could be a mistake I won't recognize for years. One moment that stands out vividly from an early LW meetup where one of the other attendees floated the idea that the Bible, despite not being a literal "Word of God", might contain legitimate deep wisdom as the consequence of a long chain of evolved tradition. I laughed at the idea at the time, and said something like "What, like 'don't murder!?', expecting her to explain with an example of the type of potential wisdom she had in mind, or else maybe laughing with me and saying "Yeah maybe not. I haven't really thought about it much". Instead though, she just seemed kinda hurt, like she was actually serious about it, and I immediately regretted laughing -- not out of mere politeness, but out of recognition that *I didn't expect that*, and if this person was serious about it then I was certainly missing something and didn't know what. Sure enough, I've found more and more things where morality is just less obvious than I thought and more fit to Christian morality than I had realized³, to the point where I'm rescuing concepts like "faith" in this sequence, and using several others in contexts where they don't provoke memetic allergic reactions⁴.

The point isn't that this claim is *true*, or that *you* should accept it. The point is that relative to everything *I* knew at the time, I was underestimating it and my current *more informed* perspective rates it more highly -- meaning that I could have more quickly gotten to results that are more correct in expectation by offering more respect to such claims -- even if you have reason to know that my current perspective rates it *too highly*.

So when the monk atop the mountain asks about the sound of one hand clapping... I mean, it's not *supposed* to be a question that has an obviously meaningful solution. It's *supposed* to be hard to figure out, and to sound like nonsense at first glance. So it's not surprising that it does. What happens if I laugh at the monk? Does he say "Oh, yeah, I guess I'm just too high and stopped making sense"? Does he get all offended at the idea that I wouldn't validate his supposed wisdom? Or does he just smile at me and look at me like I'm an idiot?

I might still laugh at the idea, like I kinda laugh at Tolle sometimes. Because I *do* think he's a bit silly, even taking into account when I've underestimated things. But I also don't think it's quite so

simple, and have been humbled enough to mostly just think "Thanks, and good luck to you, but I'm not interested" -- at least until I see something that surprises me, and suggests I pay attention.

Integration with attention

With attention, there's a spectrum where on one side I'm essentially hypnotizing people and directly dictating their experience of reality, and on the other I'm essentially going into hypnosis myself with empathy and experiencing their world directly without objection. With respect, there's a similar spectrum, with its own considerations.

When the resident homeless guy preaches to "Be the vegetable your DNA is", *that's a koan*. It's not obvious to me what it even means, so I'm in no position to evaluate whether it's good advice -- just whether I want to spare the attention to find out. When your doctor advises you to get a flu shot, it's much less koan like, but it's not *completely* un-koanlike ("what do you mean 'should', exactly?"). It's fairly rare that statements are made sufficiently concrete that you can instantly know that you've properly understood them and can reject them. If your doctor says "The flu shot will reduce your chance of getting the flu by X%", then that sounds pretty concrete, but even if you look it up and the studies seem to contradict his statement, are you sure you're looking at the right studies? Are you sure you're analyzing them properly? Is "X%" before or after conditioning on age and other factors? There is a very persistent gray area between "I am definitely understanding correctly and this is wrong" and "Maaaaybe I just don't understand as much as I think I do".

The question, therefore, isn't just "Is this objectively correct" or even "Do I have respect for this doctor", but "*How much* respect do I have for this doctor". If things were to not add up, at what point are you going to laugh him off and say "Look, I know you think you're an authority here, but what you're saying doesn't make sense, doesn't match the data, and if it did *I think I'd know*".

And at what point are you going to swallow your pride and consider whether maybe you wouldn't.

Trusting someone else's authority does imply that we're unable to look at the object level facts and figure it out for ourselves. It's foolish if we can, and fallacious if we try to use appeals to authority to position ourselves as "someone who knows", so that we can reject object level arguments instead of saying "I dunno man, that's above my pay grade". This can be humbling -- or *humiliating*, if the realization is forced on us -- but it is also often *true*, and the best we can do.

Leaning on *our own* authority does imply that we're unable to convey an understanding on the object level directly, in a way that makes sense relative to our interlocutor's models of the world. It's fallacious if the reason we can't is that they have objections we can't grapple with without softening our own stance and changing our own minds. But sometimes the inferential distance is too large to cross in itty bitty marginal little steps from a mistaken worldview, even though we can predictably take each step without encountering evidence that we're wrong. In these cases it can make a lot of sense to help people take bigger steps into the unknown-to-them, and leverage earned respect to allow strong bids for attention towards the truth, to the exclusion of what to them *looks like* truth.

Notes

1. [Fallacies are often Bayesian evidence](#), and we have finite attention. ↩
2. If I tell you that this man won't punch anyone, and you reason "Well the guy is screaming about punching people now, so evidence is against you", then you're not being careful enough because you're mistakenly counting this *against* my predictions being good despite being in line with them. ↩
3. Seeing people's response to the UHC CEO assassination makes me think "Don't murder" is actually "deep wisdom", these days. ↩
4. I'm not Christian, mind you, and this isn't a defense of Christianity as a whole. Nor do I read the bible. It was just "Oh. There's signal among that noise, and I was arrogant to dismiss that possibility. Oops". ↩

Navigating Respect: How to bid boldly, and when to humble yourself preemptively

In previous posts, we covered how [attention is how you get results](#), and why [respect is necessary in order to direct attention](#). Basically, people have to perceive a reason to listen to you. Like attention, respect can also be negotiated, and comes with its own key questions for each direction. In this post, we'll cover how to *give* people a reason to listen to you -- or notice when there isn't one and what we can do instead.

The cleanest example I have for negotiation of respect -- and the one that really clarified this dynamic for me -- involves someone telling me that she didn't think she could be hypnotized, and asking me to hypnotize her. It was a silly stance, because I had *already* hypnotized her, leaving zero chance that she "couldn't". She did have an experience *failing* to be hypnotized at a stage show since then, and I suspect she had some other reasons for making things less than perfectly straight forward, but regardless of her reasons, she was claiming that she couldn't do a thing she could clearly do *and* that she wanted to do it, and that's just silly.

So when she asked if I'd hypnotize her I said "Sure. Just lemme know when you're ready". She laughed. Okay. That's fine. Get it out of your system, and let me know when you're ready. She settled down, and said "Okay". "You ready?". She laughed again, harder. This just kept going and going.

Negotiations for respect are about the value of information. To the extent I respect you, your perspective means something to me, and to the extent that I don't it doesn't.

Imagine you've tried a certain wine and didn't like it. Then some dummy tells you that it's the *best* wine that all the *best* wine tasters like (but the rubes lack the sophisticated palette to appreciate it), and you're going to like it too. False. Empirically disproven. You tried it, and you *didn't* like it.

But what if, instead of some dummy it's actually someone who knows his wines and has proven time and time again that he aces blind taste tests and isn't lying to push a product or an ego -- and furthermore that he understands *your* tastes in particular, and is five for five on prior recommendations. When *this* person tells you that it is a bit of a weird one, but that it's actually really good and you'll learn to appreciate it if you play with it a bit... it starts to feel like maybe it's not so "empirically disproven". Maybe there's something you missed. There might be some value that his perspective points at.

The measure of this value of information is something you feel. When you start to get hungry, at first it's easy to ignore or even miss the sign that you need food. Go longer without food, and it starts to tug on your attention more strongly. If you put your hand on the hot stove -- or pick up the wrong end of a fire poker -- that burning pain is information *that you care about*, and you know you care about it by how it commands your attention. Negotiating for respect, therefore, can be expected to share this quality.

And sure enough, it does. Watch [this American Idol audition](#), for example. Any audition is a bid for attention (i.e. "listen to me sing!"), and the implicit claim is that his singing is good enough that he could conceivably win the competition. The claim isn't just something that "turns out to be false". The idea that him winning would even be conceivable is *laughable* to them. They're not just telling the contestant "No, you didn't make it", but also "We can't take your audition seriously". But learning that this piece of information is false isn't just "Oh. Okay, updated", because it *matters* to him. You don't go on a show like this without some sort of hopes and dreams attached, and what does it mean about those, if you were wrong to think you even had a chance? Does it mean you need lessons? That you shouldn't even hope? Being wrong about things like that is *uncomfortable*; even the judges were uncomfortable enough to shield their faces.

That doesn't mean it's always uncomfortable to spark laughter. If it's a joke, for example, then it's good if people don't take it seriously. If you say "Why doesn't the M&M factory hire blondes? Because they keep throwing out the Ws", then you probably don't want people saying "Makes sense. That's why I don't hire them". The difference is that those bits of laughter are rejecting bids for respect that you *yourself* reject (No, I don't *actually* think blondes are too dumb to hire. You're *supposed* to laugh at that bid for that idea to be taken seriously).

Nor do bids for respect don't have to be laughed off to be denied, though that is an illustrative extreme. If someone just ignores you, that shows a lack of respect. Or if they just turn their back and walk away, that too. It's the serious disagreement over value of information that makes things uncomfortable. There's an impulse to fill the gap, release the tension, and as a result immediately stop any respect negotiation that may be going on -- even though it may have led to productive resolution if you had let it continue.

One example that comes to mind is when I was trying to sign up with a gym membership I had bought at Costco. The lady working there told me she couldn't sign me up with that because I had previously had a membership at that gym and the Costco deal was supposedly for new members (or something like that). She then started listing details of the deals she *could* offer me, with the expectation that I accept her bid for "No Costco deal, sign up the regular way". The easy ("non awkward") thing to do in such situations is to say "Oh, okay", and either pick a deal she offered or say "Thanks anyway" and leave. Or, for that matter, to get pissy and scream at her

until security throws you out (because hey, no need to expect that to work, so it's easy to get your expectations fulfilled).

The slightly harder -- and conspicuously awkward -- thing, is to decline to accept her bid *and* decline to close the negotiation on your own. When she said she couldn't do it and started rattling off alternatives, I just thought for a moment. After recognizing that I had no compelling argument for why she had to accept the Costco thing, and that I *also* wasn't going to sign up for a regular membership on the spot, I told her "I'm not going to do that today". I could have added "Thanks, bye", and got out of the tension by cutting the negotiation prematurely. If she would have said "Okay, come back if you change your mind", I would have, because at that point it wouldn't have been premature. But I didn't know that there was nowhere else the negotiation could go, I just knew that I wasn't gonna do that today. So I said that I wasn't gonna do that today, and waited for her to respond.

Yes, it was a bit awkward. Yes, I felt the tension too. And after a couple seconds she decided that you know, she actually *can* accept the Costco deal, and just make an exception for me. I wasn't expecting that, but it was nice. Her demeanor had also completely changed from a cold and "professional" to warm, friendly, and even rapport seeking. To caricature a bit, if I had asked her where the nearest McDonalds was or some other slight favor, I would have expected something like "I'm just here for membership questions" before the shift, and "On the corner of such and such. You know the Starbucks? Yeah, right past that and to the left!" after¹.

I also found this to be useful when working with the hypnosis chat bot. When doing text hypnosis I would usually try for a phenomenon (e.g. "you feel your hand getting tingly" to start with), and then if I didn't get a confirmatory response, I would layer onto it, describing in increasing detail what the suggested experience was to feel like. I was lazy with the chat bot though, and the chat bot would just sit there indefinitely until you told it you felt it or told it you didn't -- and surprisingly enough, it was getting *better* results than I was! Looking through the logs, sometimes people just sat there for minutes, long after I would have given up and concluded "it didn't work". The "patience" of my bot held that expectation until it built up enough that they started enacting the suggestion to get out of the tension. When people *knew* they were dealing with a chat bot, the same words led to a different experience. Instead of "Oh, this person *still* expects me to respond", and feeling that tension build until it motivates a response, they knew there was no expectation on the other side of the screen. The screen was just showing words, it just felt like "Oh, I guess it didn't work", and there was no negotiation to increase "respect" as needed to achieve the results. Same words, no negotiation for respect, different results.

With my "I can't be hypnotized!" friend, it was the opposite side of this. She *knew* I was a real person, and she could *see* me holding a real expectation that she could be hypnotized if she

wants. With every laugh it was obvious that it just wasn't changing my mind -- not because I was shutting it out, just that her laughter wasn't evidence that my perspective was mistaken. It wasn't uncomfortable on *my* end, because I knew she had that capability and she wanted to be there. On her end though, for whatever reason the "I can't" had been taken pretty seriously, so when in contact with my "Oh, that's definitely not true" stance, there was some nervous laughter as we essentially played the Aumann Agreement Game². And because I stayed engaged, it continued to provide evidence that this strong expectation she somehow had of not going into hypnosis was just wrong.

It took an incredible *forty minutes* for her laughter to die out (which I was fine with, because it was fascinating to me), and for her to settle into "Okay, I'm ready". Once she did all I had to say was essentially "Okay, remember what it's like to be in hypnosis? Good. Go back into hypnosis now". After all that "I can't", it turned out to be completely *trivial* -- if you don't count as nontrivial the knowledge that she definitely could, and the ability to recognize when she was actually ready to try.

Like bidding for attention, bidding for respect can be as simple as repeating the same thing, or just waiting. The important part isn't which specific things you do to bid, because that's all downstream of embodying the belief that your perspective is meaningful. The important part is whether you know you can support the bid, because that determines what you'll find downstream.

If you bid strongly, your bid will probably be rejected -- at first. They might ignore you. They might laugh at you. *They might be right to*. Have you thought that through? Do you know what laughter means, and at which point you will start to become uncomfortable yourself as you pick up on signs that you may have over asked?

When you can hold your ground in the face of laughter and rejected bids, your bid can continue to tug and create tension that tells them "Maybe I've underestimated, and should reconsider". The natural tendency is to *avoid* being laughed at, and that has its place, but when you're confident that you can justify higher levels of respect that inclination is directly backwards. Sometimes, when done in contact with reality and when reality supports it, "being provocative" has a purpose -- or, coming at it from another angle, "Sometimes the truth is provocative, and therefore it becomes very difficult to convey the truth if you shy away from saying things which people are uncomfortable with".

This is part of what was going on in the [jacuzzi example](#). When I hinted that her own perspective framed her as ugly without invalidating the idea, that was pretty provocative. Normally, when you say something like that and the person responds "Are you calling me ugly!?", you feel a pressure to backpedal so as to avoid offense. But I knew the judgement wasn't coming from me

so it couldn't be pinned on me, and I felt pretty sure the contradiction was worth at least gesturing at. The risk of her humiliating me by pointing out how her perspective didn't actually imply what I thought seemed pretty low, so I was willing to stick my neck out and take that risk. I anticipated that once we both faced reality long enough for one of us to see what we were missing, and be forced to laugh at ourselves, it'd be her. So I held my ground with a smile on my face, and sure enough she came around to the idea that she was taking her own neuroticisms a little too seriously, and that maybe this stranger's perspective would be worth hearing.

So that's the "bid for respect" move. "How can I set myself up for unmistakable humiliation if I'm wrong?" and then do that -- if you dare³.

When you daren't, there might be good reason for that. If you aren't *completely* comfortable setting yourself up to risk humiliation, that's a hint that *you* anticipate a nontrivial chance that you're missing something important. You're always free to bid less boldly and avoid the laughter, and a good way to avoid being taken down a peg or two is to refrain from putting yourself on a pedestal to begin with. In fact, you can *work to climb down*. Maybe take yourself a little less seriously. Maybe offer to take *them* a little more seriously, and ask with sincere intent whether they're feeling respected enough by you.

This is *also* part of what was going on in the jacuzzi example.

I did poke a little fun by framing her own perspective in words that made it *visibly* silly, but I didn't *tell her* that she was being silly. I mean, maybe she wasn't. What if I'm wrong? It was clear to me that *she* wasn't seeing her concerns as silly. I didn't really feel like I could *overtly* laugh at her, while looking her in the eye and caring how she feels, and *not* feel a convincing amount of "Oof, maybe she's worried about something important that you're not seeing, and you're just being an overconfident ass". So I didn't do that, and instead offered to take her seriously. As seriously as she wanted me to, while also not hiding from the fact that it *did* seem a bit silly to me.

The words I said were "Yeah, don't do it. Your makeup might wash off, and then people would see what you really look like", but one of the things that was communicated by *the fact that I was saying it in the way that I was saying it*, was "I'm willing to take you as seriously as you want. How seriously is that?".

Partly bidding for respect by poking fun with a hint of a smirk and a wink as if to say "I bet you'll come around to my perspective, and I'm willing to eat a little humble pie if you can tell me I'm wrong". Partly giving her room to give me respect by offering her as much respect as she wanted and needed, so that she didn't need to make it a zero sum fight out of it, and we could simply play with the data as it comes out.

In this case she did come to agree with my initial perspective, but just like the attention question of "Am I missing something?", the answer you get in response to "Am I taking you seriously enough"/"Is my own perspective worth less than I think?" can be that you *are* overvaluing your own perspective with respect to theirs. If you're confident in your own stance, you can see if you can hold that smile and confidence for forty minutes of them laughing at you, and sometimes it'll pan out. In other situations though, you're going to feel your confidence dwindle at each moment their confidence doesn't, and the quicker you get out in front of the inevitable reality the better off you're going to be.

Most of the time, it's about saying "oops" earlier, rather than more boldly laughing at people or inducing them to laugh at us. The extremes serve to illustrate the full space of possibilities, and the directions in which we might want to take marginal steps, but most steps *are* on the margin. And at least at first, most of them are in the direction of humbling ourselves.

To give a more common place example, think of the stereotypical guy whining that his stereotypical girlfriend "[Doesn't want solutions](#)", and judging her as "emotional" and "irrational" for "just wanting him to *listen*"⁴. When he says "Then don't talk to Becky. Problem *solved*", she probably doesn't spend even a moment being present with that and feeling her perspective shift, because she *immediately* rejects his perspective as coming from a place of *not getting it*. And when she complains that he's not listening, he *also* probably isn't spending any measurable fraction of a second being moved by her words -- instead immediately blowing off her concerns as not-informed by the critical piece of information that he just gave: "This is a non-problem". This "You need to listen!"/"Nuh uh, you do!" problem is a common example of where disagreements fail to resolve due to missing respect. No amount of bidding seriously for attention will help, because the bid isn't *disagreed with*. It's *dismissed*.

So how can we navigate the layer of respect in order to resolve this?

Theoretically, we could just poke fun and troll our boyfriends/girlfriends until they recognize that we're actually definitely totally right, and not at all flinching from reality. And if you can pull that off, good for you. Realistically though, we're usually not so confident and usually can't pull that off without that little "I wish they'd just give me the respect that I deserve!!!" nagging through and pointing out that we don't have the security needed to play this game right now.

The option that we *do* usually have, is to notice that this lack of security (which shows up as "I can't actually have fun and anticipate that they'll validate my rightness when I poke fun at them") indicates that we don't *actually* know that we're right about our perspective being worth what we present it as worth.

And then, once we notice, it's pretty natural to consider the other person's perspective a bit more. How *could* it make sense to want to be listened to more than to have your boyfriend try to fix your problem? Oh, right. "He's an idiot that *doesn't know* what the problem is". *That's* why it's important to her that I listen first. Maybe the problem isn't just the nail, but the fact that she somehow ended up with a nail in her head, and it's likely to end more poorly next time. Seriously, who the fuck sees a nail in their girlfriend's head and has no curiosity about how it got there?

Or like... maybe he *does* get it. Maybe it really *is* just about the nail, and it was a freak accident which won't happen again. Have I actually considered that? Do I even know what it would *look like*, if my boyfriend had already done his homework, and I'm just not giving credit where it's due?

If you expect people to wonder in awe when you ask stupid questions like "What is the sound of one hand clapping?" or actually look for how you could be right when you tell them "Apparently you can just decide to not swell" -- or even just ask for your perspective on the necessity of their makeup -- you're going to invite push back at first. People are going to laugh at you, and not take you as seriously as you want, and watch to see how you respond.

And if you can stay engaged with a smile on your face as you face the challenges, you'll earn your way there.

When you can't, because *your own beliefs* are shifting in response to evidence that maybe your perspective isn't so superior as you think, then the earlier you start acting in line with your own best estimates the better off you'll be -- in your own best estimates.

The bolder you let yourself consider being, the easier it becomes to notice the calls for humility. Like target shooting and negotiating for attention, negotiating for respect is simple. Aim for the truth, and make up your mind before you flinch.

How much respect *do* you deserve here?

Notes

1. It might be tempting to wonder "Is that really *respect* though?". Maybe it was just flirting, which doesn't imply she thinks you're right about anything. Or maybe it's simply avoiding awkwardness. And the answer is "Yes, it is respect".

There are a couple things here to notice.

First, respect is expected value in attending to a thing, not probability that a thing is right. If we someday elect a president who thinks it would be a good idea to "solve poverty" by simply banning it, and you find yourself in the position of adviser to the president, what are you gonna say? "Pshh, that question is too stupid to justify with a response"? And get replaced with someone who *doesn't* respond that way? You *could*, but do you expect that to lead to the outcomes you want? Or you could say "Help me understand where you're coming from here, I want to make sure I get this right"? The latter is very obviously *more respectful*, and that does not at all imply that you think it's a good idea -- just that you recognize it's worth engaging with his ideas because he's the one that has power to do something.

Secondly, because of this, respect *generalizes*. Maybe you listen to someone because you think they have all the right answers, or because its important that you can *make sure they do* because they're the one with the finger on the button. Or maybe because they're kinda cute and you want them to like you. Or maybe you are just uncomfortable holding a firm "no" because you don't respect your *own* ability to justify a "no". Regardless of the reason, the result is the same; you're listening, and following their lead.

If you "respect your physics prof on the subject of physics but not politics" and respond to his broaching the topic of politics with "Obviously you're very intelligent and have lots to teach, but what makes you think the things you have to teach extend beyond your expertise in physics?", you're still considering his bid even though you would have otherwise not been interested in going there. On the other hand, if you were to respond "Yeah yeah whatever, no one cares what you think. Now explain this equation, physics monkey" then no one is going to confuse that with respect.

Or another way of saying it, if there's only one direction you're willing to "follow their attention" then it isn't *them* directing attention, and you are demonstrating no respect. ↩

2. "[Aumann's agreement theorem](#). roughly speaking, says that two agents acting rationally (in a certain precise sense) and with [common knowledge](#) of each other's beliefs cannot agree to disagree"

This relies on belief in the other Bayesian reasoners conclusions are meaningful evidence, which is another way of saying "respect". ↩

3. As another example of "techniques fall out of the framework, when they're appropriate", "How can I set myself up for humiliation" sounds like a good description of the iconic PUA technique called "peacocking", where they dress up in [goofy hats and goggles](#). Note that this isn't one of the *notorious* techniques, and this one gets them laughed at more than hated. I'd be embarrassed dressing up like that for no reason, and that's the point. ↩
4. Reality isn't stereotypes, of course. I actually identify with the stereotypically female frustration here more than the stereotypically male side. But that just means that we have to watch out on *both* sides. ↩

Part: Security

Security is what makes room for bids for respect.

The necessity of security for play, and play for seeing reality

Sometimes, when we suggest things that people aren't inclined to agree with, they'll laugh at us. When that's the case, we can play along and if our view doesn't change, [sooner or later theirs will](#). The way this can fail is when one side is no longer having fun with the "making fun", and begins to act on a need to defend themselves.

Other times, people skip right over making fun and go straight to making war.

I'll illustrate with an example.

Music festival conflicts

Once upon a time I was at a music festival, and like happens in places that are shoulder to shoulder crowded, I accidentally bumped into someone else. She was Ms.Grumpy pants about it, and I don't remember what she said exactly, but it was something intended to insult me and hurt my feelings. It didn't work, because "Nah, you're obviously just grumpy and unreasonable", but saying that would have escalated straight to a fight with no laughter phase. The reason this skips the respect layer and goes right to security is that she never actually believed that her actions were reasonable. She didn't have to transition from "Haha, how dumb!" to "Wait, might he be right!?". She knew she was out of line, and would have felt threatened by someone shoving that obvious fact in her face.

I didn't have anything to gain by fighting, so I just turned to her and asked "Are you okay? Will you forgive me?". It's a *little* silly, because the answers to these questions are pretty obvious. And I may have been a *little bit* laughing about it inside, and trolling the way I trolled the jacuzzi girl, but I was mostly just sincere. I really didn't mean to bother her, it really was an accident, and while I had a pretty good idea of the answer I did want to verify that she was okay and would forgive me; it is her decision, after all. And she wasn't going to make the decision without being invited to. So I invited her to, and gave her the space to decide.

She didn't seem to be expecting this, like she hadn't considered the possibility that I would *not* be hostile in return. Clearly uncomfortable with the situation, she muttered something about forgiving me, and then left me alone. That's a win.

A little bit later, in an amazing coincidence, someone else bumped into her. Again, this girl was Ms.Grumpypants about it. The girl that had bumped into her took offense to that, called her out for her anti-social behavior, and things escalated. When security came over to break things up,

the first girl put on an innocent face while the second girl remained angry (and justifiably, IMO!), so guess who got kicked out and who got a free water bottle out of the deal.

Would learning NonViolent Communication have helped?

I had been thinking about NonViolent Communication and its proper place, and this scenario makes for an interesting test. *Clearly* there was a lot of "violent communication" on both sides. *Clearly* things would have gone better if both sides -- or even either side unilaterally -- adopted the spirit of NVC. And equally clearly, teaching these women about NVC would be completely ineffective. Intention-to-Treat is not the same as Per-Protocol

Let's look at what NVC would encourage and why this encouragement wouldn't work.

The first woman, instead of saying something like "Don't bump into me, asshole", would be encouraged to say something like "I feel frustrated and uncomfortable when someone bumps into me. Could you please be mindful of my space?".

The second woman, instead of saying something like "It was an accident, no need to be a bitch about it", would be encouraged to say something like "When I hear your response, I feel defensive because I value understanding and want my intention to be recognized. I notice I accidentally made contact with you, and I need mutual respect in our interaction. Would you be open to acknowledging that this was unintentional while we both take a moment to check if we're okay?"

Can you imagine either woman being inclined to say such a thing because they went to an NVC workshop and [this guy](#) highlighted the possibility and explained why they should? Can you imagine either of them saying it *without* a snarl that betrays the true meaning of "please"?

Me neither -- and I say this as a *fan* of Mr.Rosenberg and his puppets.¹

So why not, and what can we do that *would* work?

When I put myself in those shoes, I feel like I don't have *space* to consider such things. It's like trying to do NVC with a brown bear guarding her cubs. If that thing snarls at you, are you going to lower your spear to invite her to feel comfort with you, or are you *way* too uncomfortable to be inviting in that way, white knuckling the spear, and doing fight or flight? You're not going to ask if someone could be mindful of your space, or acknowledge that you bumped them on accident if you can't handle hearing "no". If you stick to the script and say the words, your body language will make it clear that you're not really asking.

While it is definitely *possible* to face legitimate threats unflinchingly, without raising spears, [it's impressive when it happens](#), and not something you could really expect to come from an NVC

workshop no matter *what* protocol they teach. It's just too real of a problem to address with technique and choice of words. There needs to be *actual security*. The reason I was able to respond effectively in that situation is because *I wasn't threatened*. She couldn't hurt me. Whatever flaws I may have and be sensitive about, *she* wasn't finding them. The question "Don't you think you're being kinda a bitch?" would be threatening to her, because the answer kinda seemed like "yes", and what do you even do with that? How do you see yourself as "okay", if that's true about you, and you don't know how to *not* be a bitch, given the things that compelled you to act in such ways?

In my position though, it was just "No, I don't". So I had space to think, and to figure out how to give *her* space to think -- and therefore to realize that she didn't need to pick a fight.

First principles of security

Security, on a fundamental level, is an ability to compute acceptable responses to the new information we're presented with. Whether we're "okay" has nothing to do with the objective reality in isolation. [My cousin with the burned hand](#) didn't flip from "not okay" to "okay" because his injury healed in an instant. He flipped because his *expectations* changed in an instant. He went from "I don't know how to handle this!" to "Lol. Of course I do, actually".

In this light, assurance that "you are okay" is a statement both about the what the information says about reality *and* about their current state of knowledge. I could tell my cousin that he's okay *if* I'm willing to assert that the pain information has no relevance to him -- but that would be a little silly, without knowing what lesson(s) he's learned and which he hasn't. In this context, it made more sense to ask instead, and when I asked, he found right away that actually he is okay.

The wider the range of information we can face productively, and the wider the range the people we're talking to can face productively, the easier things go. It's so much easier, it's hard to put into words how big a deal this is. When you finally get to work with someone who is *secure*, it's... nice. A relief that you might not know there is to be had.

Everyone knows this stress on the other extreme of this spectrum. It's stressful having to walk on eggshells all the time, ever careful that if you ever show anything slightly threatening -- even to let slip "Ugh, this so tedious" -- you have a fight rather than a cooperative endeavor. You can still give criticism to extremely insecure people, you just have to do *all* the cognitive work to figure out how they were causally constrained and *couldn't have* done better. "No one has told you this? Okay, then it's not your fault. Let me explain what they were supposed to teach you". If you want to expect more of them, like for them to figure this stuff out themselves like you did, then you better know exactly what allowed you to figure it out and not them so that you can portray

that as not their fault too. What's less obvious is that so much of what limits us even in "normal" situations, is this walking on eggshells.

You know that whole "politeness" thing, which people take as if it's an *inherent* good? It's bullshit. Think of all the times you've self censored in service of "politeness" or "not being a total ass and starting fights for no reason". Imagine that's not necessary, *because they won't fight*. Won't *need* to fight, because they can just *look*.

"Dude, that's *stupid*"

"Oh. Huh. What am I missing?"²

Security makes everything easy

When security is in excess, bids for respect get *easy*, and you don't have to worry about offending. You get to say whatever you want, with true freedom of expression, and not only *find out* whether your thoughts hold up, but also just find out what your thoughts *are*. You get to say -- or at least *consider* saying -- "You're the *worst!* I hate *everything about you*. Including your mother", and hear "Okay I'm sorry. What should I do differently? (and should I pick a different mom?)".

When given the freedom to go over the top without concern, it becomes hard to fixate on problems. Things often come out over the top *just for fun*, showing that despite the serious complaints here *we're still having fun*. We can still engage in playful cooperation without fear of ending up getting hurt while we each try desperately to defend ourselves. Okay, maybe you can keep your mom. Just please chew more quietly, or whatever.

Insecurity is rampant in part because as a culture we don't know how to create security, value security, or even *recognize* insecurity. Before exploring how this applies to arguments, let's first examine it in context of hypnosis where it is more clear.

Feynman was successful as a hypnotic subject, despite the well known tendency for "analytical" types to struggle with it. The stereotype is for science nerd types to be so attached to "truth" that when the hypnotist suggests "You cannot open your eyes", they *reflexively* test to see if they're being bullshitted, try to open their eyes, and succeed in falsifying the suggestion. Whew! Almost believed something wrong there.

I make fun of this response, and hold Feynman up as an example of an intelligent and effective thinker doing better, but I'm not positioning myself with Feynman while making fun of the "insecure dummies". I'm making fun of *myself* here. That was 100% me when I started looking into hypnosis. I wanted to experience it first hand to see what it's like, but I wanted it to be *real*

and to not be fooled by bullshit, so when the hypnotist would say "Try to open your eyes", I would, and surprise surprise, I could.

At least, that's one way of explaining it. There's a little more to the story.

One day I was playing around with my girlfriend (now wife), having her play hypnotist while I played subject sorta as a "pop quiz", like "can you do this *at all*". I hadn't really taught her anything other than what she inadvertently picked up by osmosis, so she didn't really know what she was doing and I decided to make it easy for her -- to start with -- by kinda playing along.

She stumbled along with a somewhat awkward but not altogether wrong attempt at a hypnotic induction, and because I wasn't "really" being hypnotized, I didn't have to test whether I could actually open my eyes; I could just *act* like they weren't going to open, like I was genuinely hypnotized, and I didn't have much reason to keep a hold of "BUT I'M JUST FAKING I SWEAR".

Eventually, she got to the point where she said something like "You can no longer remember your name. What's your name?", and instead of instantly saying my name, I just said "Uh....".

"That's weird", I thought to myself. "I guess I decided to play along further than I expected to". My name was *there*, and I could reach it if I wanted, and I knew that I was *choosing* to not push back and retrieve my name... but at the same time choosing to recall my name would be against some resistance. Against the flow of things.

So another moment or two went by, and she was starting to believe more and more that she *actually* got me and that I wasn't just "playing along". It was also getting more and more *true*, as I could feel my ability to reach the intention of remembering slipping away. If I'm in a self flattering mood I'd say "She was getting too cocky, I had to show her that I still got it", which is true. But what's also true is that it *freaked me out*. Watching myself begin to lose control of my own mind -- lose awareness of where I am, mentally, and how to find my way out of it -- was a little panic inducing. A little. Not something I couldn't push through should I have had sufficient reason, but something that *definitely* guides my behavior when I'm not explicitly choosing to act against it -- which, perhaps unsurprisingly, I don't tend to find reasons to do.

Feynman didn't seem to have this panic over the idea of losing control of his own mind, the way I and other "difficult subjects" did. For him it was "if it works it works, we'll find out!", whereas I had been taking a perspective of "I need to know exactly what's going on in my own mind at all times", which made experiencing what amounts to known delusion tough for me. I'm not saying Feynman's response is "better", in general. The more sensitive you are to avoiding believing things that you can see to be wrong, the more motivation you have to avoid delusion, and no doubt that sensitivity has been useful to me in warding off poor thinking. At the same

time, I'm not saying his response was wrong for him either; he was clearly a very effective thinker, and his intellectual security and willingness to play with unusual ideas no doubt *helped* him be effective.

I've moved in that direction myself too. The more I get comfortable with how my mind works, the more I know what kind of playing around with ideas is *safe*, and therefore not freaky. With competence comes confidence and all that, so the better I understand how to keep my thinking from being led off the rails the more secure I feel to explore increasingly different modes of thought.

Which is where "regular conversations" come in.

Everything is threatening, because we keep overextending

"Hey, have you considered [the other side]" doesn't *sound* like some crazy dangerous labyrinth to avoid, but it very much can be. Because many people much of the time³ cannot ground their beliefs.

Even when you get to something quite simple like "Is the earth flat or round?", when you try to have regular people argue against flat earthers they're liable to say extremely stupid stuff like "When you use a telescope to zoom in on a ship that you can't see anymore because it has gone over the horizon, you can see it again". The atmospheric lensing thing he's referring to does exist⁴... sometimes. But if the light is blocked by the curvature of the earth, pulling out a telescope doesn't magically change the refraction of light in the distance, so wtf? Are we to believe this guy is too stupid to be able to think one step ahead, and notice the validity in the flat earther's argument here?

Or might it be that he's *choosing* to not think things through, because he knows that it's likely to end up with "Good point. I don't know how to explain that"?

How do *you* feel telling the obnoxious flat earther that he has a good point, and that you don't know how to refute it? How do you think this round earther feels?

The information is "He made a good point, and I don't know how to explain why he's wrong", so what does *that* mean? That the earth is flat? That's a shameful belief to have in today's society; we definitely want to avoid coming to *that* conclusion. Or that he's dumber than a flat earther? That's not any more palatable. So *of course* he's going to flinch away from engagement, afraid of where engaging could lead.

This may be especially clear in the case of flat vs round earth, but it also explains essentially any debate about politics or religion ever. Especially when you have a professional political arguer show up at a college campus and make fools out of unprepared kids who don't share his belief.

Of course they can't argue their points, and if they were foolish enough to truly engage they'd have no choice but to change their minds -- and risk being wrong, and/or socially punished for it -- or, you know, to find humility and say "I dunno". But that's an option they probably haven't seen role modeled much in an obviously winning way, so that option probably isn't on their table because probably no one has put it there.

Predictable prediction error requires insecurity

Diagnosing insecurity becomes simple, once you understand what it is and what it serves to do.

If people are listening, and having fun, then updating is happening. If you make a bid for attention, and it isn't granted, laughed at, or ignored, that's insecurity. You've managed to find something they don't trust they can defend.

That means if someone makes a bid for *your* attention, and you don't grant it, find it *genuinely* funny, or *genuinely* uninteresting, *that's insecurity*; they've found something *you* don't trust you can defend.

If you're annoyed, that's insecurity.

If you're angry, that's insecurity.⁵

If you're frustrated, insecurity.

If you're offended, insecurity.

If you're sneering, that's insecurity.

If you're laughing, and think it's genuine, but you can't do it with the kind of warmth, openness, and eye to eye contact that leaves you actually feeling them being tugged into security, *you're wrong, that's insecurity*

At least, if these feelings last for more than a moment without resolution.

What these responses all have in common is that they involve being *displeased* at reality being what it is. It's a divergence between your [expectation/intention/setpoint](#) and measured reality; it's prediction error. If you can control reality to conform to your expectations, then great! If you touch a hot stove, by all means flinch away. If you're aiming at a target and drift slightly left, then sure, drift slightly right to compensate.

If reality *continues* to violate your expectations in a predictable direction, facing this quickly teaches you that your expectations are wrong and to adjust your expectations to match. Sure, maybe you're just wrong about them not being a dumb jerk and it's totally *them* that's in the wrong. But your expectation that they wouldn't be a "dumb jerk" is looking wrong too. *Clearly*

wrong, in many cases. Which means that you're navigating based on a false assumption and that's why it's so uncomfortable; you're beating your head into a brick wall of a dead end, you dummy. And you'll notice that, and update -- unless you take steps to make sure you don't.

i.e. insecurity.

Even subtle insecurity changes the game completely

This can be *really* subtle. Even when you're feeling "confident", that very well might be insecurity if it's conspicuous -- because if you really know how to face people challenging you, why the need to hint that you're ready for challenges? Why not just face them should they arise?

Even subtle effects can accumulate, because "insecurity limited" isn't a binary thing where either you "are" or "are not". Maybe someone shows you a study that challenges your preconceived notions, but the p-value is marginal it's not super well controlled, so having "analyzed the evidence", you dismiss it. Or maybe most of the studies on "this sort of thing" are negative. You might pat yourself on the back for being willing to update if the study were strong evidence, and you might be right that you would, and this might bound your irrationality. But would you have predicted that this study would have turned out this way? Are you looking at the data, and letting your beliefs shift as they do? Or are you coming up with a reason to prevent this shift -- in the same way that you naturally shield from ["When I scratch my nose you will find it hilarious and have no idea why, because you won't remember this conversation"](#)?

What do you think the cumulative effect is of passing everything through a subtle insecurity filter, where you'll update and face the consequences -- but only if the discomfort of ignoring the data [rises above some threshold](#)? And only if it *also* stays below another threshold where you can dismiss the person for being a "pushy jerk" trying to get you to read something you disagree with. Errors compound.

If you want to see how subtle this can be, watch [Dr.Mike talk to Dr.K about ayurvedic medicine](#). I would describe Dr.Mike as intelligent and open minded. He's even open minded enough to hear someone tell him to his face that he doesn't have an open mind, and nod along listening before voicing his disagreement. "Insecure" and "closed minded" aren't high on the list of descriptors that come to mind. He's clearly interested in learning about ways of thinking different from his own, and he clearly respects Dr.K on this topic

And yet, in his conversation with Dr.K on ayurvedic medicine Dr.Mike was clearly (and self admittedly) frustrated, and despite being friendly respectful and good faith, the conversation was *also* antagonistic to the point that it interfered with efficient learning. [Dr.K even stops and addressed this explicitly](#), pointing out a lot of what I'm saying here. Dr.Mike is already

functioning at a high level of openness to challenging ideas, it's just that Dr.K's ability to see what is going on with people and [compute good responses](#) is on another level.⁶

Even the not-entirely-[straw Vulcan](#) trope is a result of insecurity. Those of us with [aberrant needs for precision](#) want to be Less Wrong. This isn't entirely a bad thing; sensitive instruments can detect even minute errors, and high gain controllers control *really well* -- that is, when they're able to keep the error signal within the smaller acceptable window in which they can avoid saturation. Autistic tendencies lead to poorer ability to read faces not because it's harder to perceive the information people wear on their faces but because it's *easier*. Avoidance of eye contact comes because the information conveyed by eyes can be overwhelming, and hard to parse confidently. Math is nice partly because it's useful, but also partly because we can actually expect to get things provably right, which is a relief from prediction error. The ability to decouple enough to give calm considered analyses of problems is *great*. Yet it's decoupled, and knowing which answer we're *supposed* to reach only helps to the extent that we can recouple to object level reality, reach the answer, and believe what evidence suggests to be true.

On top of that, with few exceptions we are *meta*-insecure. The idea that you have to be perfectly clean, or you don't get to "have an opinion" in the only way you know how, and therefore don't get to feel like you're worth anything, is quite threatening. Especially to people who think they're doing the work already, and have earned it, and are now being told what sounds like "Sorry, you're just as bad as everyone else. Your efforts are not valid and not appreciated. Have some humility, loser".

So if you're *mildly annoyed* when someone doesn't agree with you, and it seems like the reason is that they're probably flinching away from reality themselves, it's *super easy* to let yourself off the hook with "I'm basically right". At the same time, that difference between "basically right" and "right" becomes the difference between someone putting up resistance to your ideas, and the person just listening because your stance represents truth. It also frequently becomes the difference between "wrong in surprising but avoidable ways" and "actually right".⁷

If you just want to rant and whine in some cases, that's entirely understandable. I don't do zero of that myself. Sometimes that's an easy way to express thoughts and emotions which feel in need of expression, and then reflecting on this can shift your perspective to better align with reality. I fear we often miss this reflection step, because we don't recognize the existence and importance of subtle signs of our own insecurity. The moment we notice, we can ask: what facet of reality are we flinching away from in order to maintain our unfulfilled expectation of success? What happens *if we look*?

No security by fiat

Insecurity is everywhere, and super important, but we can't just print out more security and have it work by fiat. It requires actual cognitive labor to parse mountains of information such that we can trust ourselves to be less wrong in more contact with reality than in less contact with reality. Those women at the music festival can't "just get along" because they literally can't expect to compute a solution by opening themselves to contact with reality.

If the first girl said "Could you please be mindful of my space?", with spear and shield down, she'd be opening herself the other girl laughing in her face. "'Your space'? This is a music festival hun!". If the second girl said "I need mutual respect" and "Would you be open to acknowledging", that opens her to a "No". And neither girl knew how to make sure this wouldn't happen, or end up better off than fighting if it did. If you try to *force* them to use NVC, you'll be trying to force them to do things that work *worse* for them in their expectation. The very fact that you have to force them in the first place shows that this prediction error is itself predictable -- i.e. insecurity. It's insecurity about "What if people don't act secure!?!?"

Security isn't an unlimited resource and it often pays to do the cognitive labor to develop more, but it also pays to keep track of when it's in short supply, and the limiting factor.

At this point, it makes sense to introduce an analogy I find quite useful when thinking about security and maintaining cooperation over conflict.

The spectrum between invitation and threat

Everything we do is a combination of an invitation to cooperate and a threat of violence.

Consider a few things you could say to someone that is doing something that you don't like.

- When you do X, it makes me sad.
- When you do X, it makes me sad. Please stop.
- When you do X, it makes me sad. Stop.
- X is bad. Stop.

The first almost looks like they're just letting you know so that you can cooperate better, and the last sounds kinda like an order that doesn't even take into account your needs - and so people are going to feel more compelled to cooperate with the first than the last. It's sorta a blend of "invitation" and "threat", with a balance shifting towards "threat" as the list goes on. Yet these aren't even the extremes of the spectrum. Even the first puts blame for their emotions on the other person ("makes

me"), and even the last doesn't declare "I don't care if you need it to live. If you don't stop I'll stab out your eyeballs".

There's a continuum here. Just about everything we say and do has at least a hint of both, and the idea is to focus on attenuating the threat of punishment and amplifying the invitation to cooperate. Instead of saying "I want you to pass the ketchup" with a subtext of "And if you don't, you're breaking the norms and being a jerk" -- which isn't far removed from a threat of punishment.⁸ You could also say it with the subtext of "Because that would be cool, and I'd do the same for you". Sure, maybe refusing to pass the ketchup is a jerk thing to do in context. Passing the ketchup would also be cool, and I'd do it for you. The degree to which we focus on one vs the other makes a big difference in the extent to which others focus on responding to one vs the other.

NVC literature will have more to say about the cleanest possible ways to word things, but the basic idea is that if you want to foster cooperation and non-violence, it helps to limit yourself to sharing your actual experience and allow them the room and trust to respond cooperatively to that without trying to "push" or "pressure" them in any way. Not "It 'makes' me scared", because why are you blaming them, if not to push them to do something about it. Not "When you are a jerk to me" if they might use different (and more charitable) abstractions to describe their behavior. Boil it down to the actual experiences. When you raise your voice, I feel afraid.

Try that on:

When you raise your voice, I feel afraid.

Notice how that has some impact, and in a different way than "Why are you yelling at me!?"?

It's there as a thing to be dealt with. It can still provoke preemptive defense like "That's not my fault!", but at least the communication itself isn't saying they did, and that outburst is essentially caused by a "delayed neutron".

NVC as neutron moderation

I like to think of nonviolent communication as neutron moderation between two masses of fissile material. Threats of violence tend to provoke threats of violence. As long as the average number of violent neutrons kicked off for every violent neutron is below one, peace can be stable. The moment it goes supercritical, you got a *problem*. With all the myriad of ways to perceive a Defect when Cooperation was intended, it's no wonder Tit for Tat with forgiveness beats pure tit for tat.

NVC, as a "technique to try" or a "ethos to [attempt to] live by" can make the possibility of nonviolent response and the costs of violent response salient, and this can moderate neutrons *on*

the margin. If you're anywhere near criticality, this can be a game changer. If you're able to notice that you were starting to spiral into violence when you could instead choose to de-emphasize threat and emphasize invitation to cooperate, then that can be the difference between getting along and melting down. And if the NVC framework can guide you by giving you an example of how it can look to communicate nonviolently, *great*.

Those two women though? They were supercritical no matter what. The only realistic option in the moment is separation -- either physically or metaphorically. "No politics at the dinner table". Good luck getting them to ask nicely and accept no for an answer.

On the other extreme, I didn't follow the NVC protocol either. "You okay? Forgive me?" is nonviolent⁹. But doesn't fit the NVC framework of "observation, feeling, needs, request". It's just that in this case I knew what I didn't have to say, and had no need to advocate for my side, and it would have been goofy and out of place to go on a long monologue like:

I realize I bumped into you, and it seems that upset you. Are you feeling angry or frustrated because you need more respect and awareness of your personal space? Would it help if I'm more mindful of giving you space in the future?"

If the system isn't near criticality it's just not near criticality, and it's not hard to come up with a nonviolent response.

As a result, efforts to do NVC can be anywhere from woefully insufficient, to absolutely critical, to just not needed.

Or, as we'll cover in the upcoming "navigation" section, actively counterproductive.

Sometimes we have to slash and burn around the edges to create a larger field for safe play.

In the next post, we will cover how to get your two year old daughter(s) to enjoy getting her flu shots. The challenge I leave you with today has two parts. First, How do we do that? Second, how is enjoying flu shots rational?". I have ended on a clue.

Notes

1. [His workshop is available on YouTube](#), and is pretty good ↩
2. "Sorry for being a overly mean about it", "No problem, idiot".

The prevailing norm that the initial aggressor is *unconditionally* deserving of scolding is mistaken. If the communication channel can handle such things in play, then there *is no harm*. ↩

3. Most people most of the time? ↩

4. [https://en.wikipedia.org/wiki/Fata_Morgana_\(mirage\)](https://en.wikipedia.org/wiki/Fata_Morgana_(mirage)) ↩

5. In [this clip](#) you can watch the kid move from subtle insecurity -- and getting challenged because of it -- to flat out stating the truth.

Notice how much harder it is to take a step forward when the kid is [no longer telling you that he's not sure you'll take him seriously?](#) Notice how much less wiggle room there is than if the kid were overtly *angry*? ↩

6. I'm [not alone in this view](#). People make "[How to flip someone from rude to respectful](#)" videos studying his conversations.

If you want to develop a sense for what it looks like to intentionally navigate insecurity and skillfully face surprising and challenging information that comes up in conversation, Dr.K's conversations with people are a great place to start. ↩

7. A lot of this sequence has been about what looks more like "psychological issues" than "rationality". For example, while one might describe makeup girl's idea that she couldn't risk her makeup washing off as "irrational", it's the kind of "irrational" that looks like "a psychological issue". It's the kind of thing for which you'd say "You should see a therapist about that", not the kind of thing for which you'd say "You should read LessWrong!". It's still about her not seeing reality for what it is, but the cause of that looks less like "Figuring out what is true can be technically challenging" and more like "She's not even trying to make sense and doesn't even believe herself because something wrong in her head".

When insecurity gets *more subtle* though, people actually *do* start believing their own rationalizations of insecurities. When people say "That study really is bad! I really *do* update on evidence!" and "Being mildly annoyed doesn't mean I'm insecure, I'm annoyed with what *he* is doing wrong!", they often *actually believe it*. When these insecurities are subtle enough that they are legitimately unseen, it goes beyond "solving psychological issues" and becomes *directly* about the practice of rationality itself. ↩

8. NVC isn't doing the oversensitive "Words can hurt *feelings*, which is *basically* the same as violence (if you've never experienced the latter)" thing. It's doing the libertarian "All law is ultimately enforced at gunpoint" thing. Break enough little norms, and you'll be forced to concede or break bigger ones. If you continue down the path of "violent" language for long enough, you'll learn the reason for the naming.

The language that NVC refers to as "violent" is language that threatens and leads towards real *sticks and stones, break your bones* violence. It's just a question of how far that path is taken before someone backs down or disengages ↩

9. When insecurity is threatening to explode everything and you're wanting to play it safe, then you are going to be very careful to strip your behavior of all implicit threats and focus entirely on direct observables that they can't take issue with... which sounds a whole lot like NonViolent Communication ↩

Navigating Security: Fighting flammability with fire (when safe)

So insecurity is *the* driver of both irrationality and psychological "stuckness". It's why we can't "just look at reality" and instead get ourselves wrapped up in both extremely subtle and sometimes not-so-subtle flinches from reality that necessitates a place like LessWrong in the first place¹. It's nice to have a concrete and actionable picture of [how we can earn respect](#) and [coordinate attention](#) towards the right answer conditional on security, because that helps us notice when we're insecure. And it's nice knowing that [insecurity is the potentially rational perception of our own inability to handle information productively](#), because that illuminates a model of secure behavior to strive towards, without getting caught up in nearby confusions like "insecurity is when we think maybe bad things".

But we still need to know what we can actually *do* about it when we notice that insecurity is messing up our rational thinking and rational discourse. We still need to know what it is that we can do when someone is flinching into insecurity, that will cause them to instead turn towards reality and manage it. Including, of course, when that person flinching is ourselves.

The first thing that comes to mind, which follows straight from the definition of insecurity, is to [create lines of retreat](#). If we flinch because we know we can't handle the truth, then maybe we should put in the work to create that ability by figuring out what we *could* do, *if* we were to come into contact with the information we're terrified of hearing. "What if my makeup washes off and they think I'm ugly!?". Good question. What if? These questions actually do have answers, and once you're comfortable answering the question, you will no longer feel the need to flinch from asking the question either.²

While there is much to be said about how to put in the work to construct security, that's a bit of an expensive last resort, and we have not yet exhausted our options. Sometimes it really just feels like "Aw, c'mon! You can handle it!". What if we're pretty sure that's *true*, but the other person³ disagrees? How can we credibly 'bid' for security such that our bids are compelling, and we get to skip re-doing the work that has already been done? How can we *create* security from known-truth, and do it *as efficiently as possible* so that we can get back to "just pointing at what's true"?

Just like with attention and respect, bidding for security means being comfortable with the risks of going too far so that you can [set your sights on the target](#). In this case, it means doing things

that are *scary* -- perhaps to yourself -- yet knowably and demonstrably safe.

Taking my little girls to get their flu shots was a fun example of this. When my first was two, we used to play "silly or scary" a lot. The win condition for me is to scare her *just enough* that she says "Scary!" -- meaning "Enough, no more!". The win condition for her is to say "Silly!" -- meaning "You don't scare me, I dare you to do worse". On its face, this looks a lot like "exposure therapy", and it is. The focus is different though. I'm not trying to extinguish fears, and my goal isn't to help her be less afraid over time. The proximal goal is to *have fun*, and specifically, to have a special sadistic kind of fun which is quite adaptive. Because it's truth tracking.

Our world is complex. Sometimes things that seem good are bad, and sometimes things that seem bad are good. The phrase "It's like ripping off a band aid" points at this phenomenon and our screwed up way of dealing with reality. Good things should feel good, bad things should feel bad. If a thing that's *good* to do, like taking off a band aid, *feels bad* to the point where we try to avoid the sensation by getting it over with quickly, then we're out of touch with reality. What would it like to be more in touch? What would it be like to *notice* that removing a band aid isn't going to harm you?

Or, to ask one of the goofiest questions ever:

What's the *rational* way to take off a band-aid?

The answer, is that it would be like re-contextualizing pain. Feeling that sensation which we associate with badness -- in full scariness -- and realizing that it's not actually a problem. It's actually okay, and maybe even a good thing.

This might sound weird, but it's something we already do. Some of us enjoy spicy foods, and don't interpret that burning sensation as cause for panic -- at least, until it gets stronger than we're used to. I think almost everyone understands what it means for a massage to "hurt so good". It's "pain" sensations, yes, but the *meaning* of these sensations is relief, and it's *safe*. Similarly, roller coasters, jump rock. Fear, combined with perceived safety, becomes excitement. Sad movies, and sad music, can be beautiful. And we don't flinch away from them because when we're secure we notice the difference between the message and the messenger. Reality is not our base level sensations, nor is it the mistaken abstractions built atop them.

Taking off a band-aid stimulates our nociceptors a little bit. It "hurts", but it doesn't *hurt* us. Taking off a band-aid "rationally" means noticing this. To the extent that these sensations are bothersome enough to try to get away from, we have something to learn. A little masochism is rational here. Teaching people to relate to band aid removal well requires putting oneself in their

shoes, showing that you're in it *with them*, and then demonstrating that *this feeling is okay*. It's empathetically reflected masochism for them to follow. In other words, empathetic sadism.

Like band-aid removal, flu shots are another one of those things that on the surface seem hurtful and scary, but which a more grown up understanding shows to be actually a good thing⁴. So how do you teach this to your two year old daughter such that she has fun, and doesn't feel a need to flinch away from the harmless reality?

You bid for security. You ask yourself the relevant question, "Am I willing to risk someone getting hurt?", and find that of course you are. Flu shots are safe and worthwhile. The bigger fear is "What if she gets scared?", and the answer is...

Then she gets scared.

That's *fine*.

"Psychological trauma" is scary and harmful if you don't know how to repair it. The solution, as with everything else, is "look at the truth". Is she okay? Yes? Great, look at that. That's what is needed. That is *how* you "repair trauma".

I'm not just saying this from a theoretical perspective, or dismissing it without empathy like "Nah, she's fine" while someone else is suffering and I'm not. I've been there. I'll spare you the gory details, but I went through a very bad experience which left me with PTSD symptoms for a while. Specifically, I'd get bouts of seemingly uncaused anxiety just out of nowhere. The situation *as a whole* wasn't at all easy to deal with, because the answer to "Am I okay?" was "No. Not really". I still had the very real problem of figuring out wtf to do with the unpleasant information I was presented with, and that was probably the hardest thing I've ever had to do. But to the extent that I *was* okay? Like, whatever, the anxiety doesn't represent an ongoing threat, so who cares. It's fine. That part wasn't a problem. It was just "Ooh, feelings which obviously don't map reality. Neat! Anyway..". Over the course of a month or so they stopped coming up. The hard part isn't the feelings, it's when the feelings seem to be pointing at something *real* and *important*, and we *don't know what to do about it*.

So, what if she's scared because the shot seems scary and bad to her, and it's not? What if you can't just tell her "It won't even hurt, it's no big deal really" because she won't believe you and it escalates to insecurity before she will? Fine, we're doing it anyway, lol. Put on your evil grin and be the sadist. Enjoy her pain, because *you* know what to do with it, and *you* know that she's fine. And because you know that when you invite her to get in on the fun, *she'll feel the tug*.

It's not easy enjoying the pain of those you love

Benevolent sadism isn't always an easy thing to do, and it takes some diligence to catch yourself worrying about feelings and refocus on the underlying reality⁵. A woman I care about texted me once about how she ate too much weed brownies before her flight and was kinda freaking out due to the unusually intense turbulence. I don't like seeing her distressed for no reason, and my first impulse was to "help".

But would telling her that she's safe have the desired impact? Not really. Some, but not enough. Would she laugh at me, because she honestly thinks my perspective isn't very meaningful here? Definitely not. Was it a security thing? Yeah. Was she in real danger? Nope, definitely not. She *is* okay. Might not *feel* okay, but *is* okay, and that's what matters.

So my job, I had to remind myself, was to enjoy her suffering *for her*. And do it in such a way that she feels invited to join me in my sadomasochism. So I teased her. "Well, it was good knowing you. Too bad this is how you have to go". "Oh, yeah, I know. You're not joking and might have an actual panic attack. Can you send me video please? I wanna see.". "Well *I* am having having a good time. And you like when *I* enjoy myself, don't you?" This can be the hardest part to communicate, because unless you've been there -- on one side or another -- it really just sounds like "Oh, so you were an *asshole*".

Which isn't exactly *wrong*, but there's a big difference between the kind of "asshole" that gets a smile along with the accusation and the kind that doesn't -- even though it can be tough to tell from the outside at times. Sometimes so hard that your best friend from second grade can't tell.

Security is created by proving that egg shells needn't be walked on

One time I was out camping with some friends and family, and after we got back my friend asked me why I was such a jerk to my cousin's girlfriend. At first I had *no idea* what he was talking about, but he explained that he was talking about when we were getting ready to go somewhere and I was saying stuff like "Jesus Christ, you're taking forever. Get your shit together, woman" -- and it didn't sound like I was joking. Which was funny, because that's *exactly* how I remembered it too, and yeah. That sounds bad.

But it came off differently to me -- and to her. It was just sufficiently clear that I wouldn't actually be annoyed with her for taking too long -- or rather, to any extent I might actually be annoyed I'm not *endorsing* the annoyance. I'm not *attached* to annoyance. I'm purposely exaggerating, and therefore communicating that I will laugh it off if she just told me to fuck off⁶. When I confirmed with her that this was indeed her read of the situation⁷, she told me that she

would have found it *weird* if I were to say "Jk jk!". Like "Why are you so insecure about this? What are you afraid of, that you're not admitting?"

The reason it would have been so unsettling and out of place is that I had previously teased her about things she's much more sensitive about -- like being a "diva" when genuinely upset and arguing with my cousin. Doing this *actively created* more room to play with potentially upsetting things by proving that even when she's genuinely upset and not showing her best self, and I'm calling her out on it... I actually thought she was pretty cool still, and worth him keeping around. And if that's safe, then how can you perceive danger with something as silly as "Get your shit together woman", when she's taking a few extra seconds, and I don't even mean it? If I were to signal that *I* was afraid, by blurting out "jk jk!"... do you see how that's evidence *against* security, and why it would have been disconcerting to her given the context?

When things are secure, you prove it by doing security, rather than coddling when coddling isn't necessary. You leave off the "jk jk", so that people get the chance to learn that it goes without saying. You call people "diva" rather than asking if they forgive you, when they're upset about silly things. You nudge people *towards* breaking down into tears, because when things are secure it is safe to play. And the further you can play the more you can learn, and as a result the more reward you can reap.

She was explicitly grateful that I didn't congratulate her for getting on the dirt bike at all, and instead pushed her to climb the hill the boys were riding. After she fell twice, I told her "Don't be a bitch, get back on the dirt bike and get up that hill". The reason it didn't come off as 'mean' is that she knew I was with her in her shoes as I said it. If she were to break down and cry after a third fall, I wouldn't judge her negatively for it. Pushing herself to the point of tears was her *choice* -- and one borne out of strength, not weakness -- and when you recognize that, the option of strength feels more compelling. She knew she would have space to figure out what to do with her new information, and that at the end of the day she would come out ahead in one way or another. So she took the risk, fell, and shed some tears. And then got back on the bike for a fourth time, and made it up the hill.

This is what I mean when I say that excessive "neutron moderation" can be actively counterproductive. If things really are secure, then why advertise insecurity by saying "jk jk!" when goes without saying? Like in a power plant, flirting with criticality can be useful and bring warmth when done in controlled fashion. It shows you where the boundaries of safe play are, so you can fully *relax* -- knowing not only which areas safe to play in, but also trusting that *should you find the limit*, you can bounce off and recover.

When it came time to talk to her about emotionally scary "losing the man you love" stuff, it was sure nice being able to just tell her that what she was doing was *dumb*. It's *really nice* not having

to pussyfoot around the obviousness of the mistake, or waste time validating what ultimately amounts to known-incorrect ways of looking at the situation. No "But you have to worry about her *feelings!*" and no "But that doesn't make sense!". I didn't have to spend time explaining why my somewhat unintuitive advice of "You don't have to respect his decision to dump you" was actually correct, in this instance. She figured that on her own shortly after entertaining the idea, and he took her back in their next conversation. They're now happily married four kids in, so I don't think it's premature to say "I told you so".⁸

So "sadistic asshole", yes, but for a *reason*.

You'll know you're doing it right when they tell you

As I mentioned earlier, there is a big difference between the kind of "asshole" that gets a smile along with the accusation and the kind that doesn't. If you're doing it right the feedback you get should sound like that given to Frank Farrelly as described in his book "Provocative Therapy"⁹

(Sincerely, warmly.): You're the kindest, most understanding man I ever met in my entire life - (Grinning) wrapped up in the biggest son of a bitch I ever met. (T. and C. laugh together.).

My friend, during her flight said something similar to me: "You always know how to make me feel better. Though if I told anyone else that, they'd be picturing something *very different*". Her answer to that last questions was yes, by the way. The fact that I was enjoying myself *did* make the experience -- fear and all -- enjoyable for her. She seemed a little mindfucked and unable to make sense of the seeming contradiction between suffering her fear and enjoying the fact that she was suffering in fear, but her honest answer was "Yes".

I'm not interested in "extinguishing fears". I'm not interested in shielding people from feeling things that are real. What I *am* interested in, is *capacity to feel and still function*. The important part is capacity to look at reality, without flinching, and *play with what is real*. Once you have that, noticing the reality itself becomes easy.

Test your intuition, spot the fake

So what'd I end up doing with my daughters and their flu shots?

I'll tell you two versions of the story, one that accurately represents what happened and the other which is BS. See if you can tell which *doesn't feel right*.

1) Despite what I've been emphasizing here, I obviously recognize the limits of what one can expect from a two year old. You're literally stabbing a piece of steel into them and no matter how you look at it, *it does hurt*. No one likes that, and it's just inherently scary. I don't actually want my daughter to be traumatized, so I explained to her the importance of *being strong*. I assured her that it was only going to hurt a *little* bit, and that she would be *fine*. I reminded her that it's important to trust adults over her feelings because us adults know better and feelings are often irrational. The nurse backed me up, saying "It will quick, I PROMISE! And then you get a band-aid!". As a result, despite starting out a little concerned, the nurse and I were able to change her mind; my little girl smiled the whole time and wasn't the least bit scared. She told me "It was easy! Fear is irrational!".

2) I told her that it was *scary and bad*. So scary and bad that even Dada was frightened. I purposely teased her, to see if I could get her to be afraid. When I succeeded, I teased her "Neener neener, you're afraid!" and also "As you should be. It's *definitely totally* scary, you know. I'm scared too!" She couldn't quite tell if I was scared or not, or if she should be scared or not, but she could tell I wasn't afraid of the fear. She could tell that I was *playing* with her, and daring her to play back and prove that she's braver than I was giving her credit for. When I exclaimed that it was going to HURT, the nurse panicked a bit and asked "What are you doing!?", but my kid didn't take my overly dramatic whining very seriously. She had fun playing "Braver than Dada", and the whole experience went smoothly with no tears... until we told her that she couldn't get another shot until next year, at which point she *melted down*.

Can you tell which is which?

Okay, I think it's obvious that the first story can't be real.

It was kinda funny having to deal with a tantrum *anyway*, as if the universe was teasing me for getting cocky about it. Fine, I asked for that humble pie so I'll eat it. When my younger daughter was two, I dialed it down a bit. She started to cry for less than a second, but I asked her a few seconds later and she said she was happy to get another. This seems to have stuck, because next time she was dragged to a doctors appointment, she started excitedly talking about how she was going to get a band aid, and how "it didn't hurt!". Thankfully, she took that let down well too.

I wanted to tell the fake story too to show how obviously fake it is. Even when I don't explicitly say "and then everyone clapped", the wishful thinking vibes stand out. When you have an adult clearly worried about something, trying to tell the kid to be less worried than they are, there's no way you can look at that situation and actually expect it to work. You can claim that it's the "right" way to do things, argue that it's "evidence based" (whether it is or not), but you can't say

"This will work, and the two year old won't be afraid". Or if you do, you'll be lying because when it doesn't *you won't experience surprise*.

The real story is weirder in that the setup is less common, but the conclusion actually follows. When you show kids that it's not safe to feel, and that you're freaked out, *of course* they're going to learn that it's not safe to feel, and are going to be freaked out¹⁰. When her feelings are *okay* to you, and you prove it by playing with them even when she's scared, then your comfort shows her that she can be comfortable too.

Shut up and do security, find out if things are secure, or notice that they're not

If you want to bid for security, you have to face the threat that provokes insecurity, and remain secure. So your little girl is going to be scared of the shot. Can you handle that, and show her what is and what is not worth fearing? So some girl at the pool might feel insecure about her makeup washing off. Is that okay *to you*? What if she accuses you of calling her ugly? What if your friend's girlfriend is being kinda a diva? Can you poke *fun* at her without poking *malice* and trying to "should" her into behaving differently, as if her current behavior isn't *okay*? If she were to lash out and insult *you* over it, would *you* be okay, or would you have to raise your hackles and your defensive shield?¹¹

What if they get hurt?

What if they hurt *me*?

What if feeling insecure is *correct*?

When my cousin [burned his hand on the fire poker](#), I *knew* pain couldn't be the issue, but I *didn't* know whether he had learned everything he needed to learn from this little bit of uncomfortable new information. I didn't want to push him to throw out that information for *no reason* (or worse, do it because *I* am uncomfortable seeing him in pain!), so I didn't. I went the other way, towards "Are you okay?" -- or rather "[What is it exactly that's not okay? Is it that it] Hurts?"

This is the same kind of move as "Am I missing anything?" and "Am I taking you seriously enough?". Rather than pushing my own bids, I'm accepting his, and in doing so, demonstrating that there's no need to push his side to the exclusion of hearing mine.

He feels not okay. Okay, maybe he's not. What's that about? Turns out that when you ask if you're missing anything, it tends to lead to you missing fewer things -- including in the eyes of the other side of things. When you ask about what's not okay, the same thing happens.

Similarly, [at the music festival](#), I have no idea what was going on with Ms.Grumpy pants. I can't tell her that she's okay, because *I* don't know what kind of mess of unintegrated information she's

dealing with. Even if I did, scaring her more would lead to her fighting back and I didn't want to get kicked out for foolishly attempting to help someone ineffectively and without appreciation. So no, not gonna tell her that she's okay. I can ask though, and giving her that space to examine gave her a little more room to realize that she was *okay enough* to drop her beef with me and go back to half enjoying the show.

Sometimes it's worse than that though, and even "Are you okay" doesn't get more security because the answer is "NO!". In these cases, there's not much we can do to expand the width of the communication channel, and we must make do with what we have.

One example that stands out because it took me to my limits involves a very drunk friend of a friend that I had just met at a party that night.

When things *aren't* secure

He was having a meltdown over something or other, and was threatening to drive away which would have been *seriously* unsafe in the state he was in. People were telling him "Dude, that's *stupid*", and while I'm a fan of that approach when it works, this wasn't a case where it was gonna work. Neither was "C'mon man, don't do that", which people were trying. Just like the fire poker incident, I had *no idea* how to proceed. "Yeah dude, don't drive" sure seemed right, but also not effective. It'd be instant failure of security, just like every other time people tried it in the preceding seconds.

Normally, the question I use to point at the direction here is "Are you okay?". In this case though, he *clearly* wasn't. He was beyond "not okay", in some really fucked up place, for whatever reason. Can't get a bid for attention through, or respect, *or security*. I don't know *what* the heck is going on here, but I know that if *I* were so dismissive about the very high risk of a serious car crash, whatever it was would have to be *bad*. Taking a moment to ask myself that question "What would it have to be like for *me* to respond this way" was sobering. It was a "Oh *fuck*" kind of moment, because it would have to be really, *terrifyingly*, bad.

Not knowing what was going on, I didn't know what to do and didn't know how to expect "okay" outcomes with any confidence at all. So bidding for security by deliberately adding *more* uncomfortable emotions to his plate was off the table.

So I didn't bid. Because things *weren't* okay.

In other words, there was some extraordinarily threatening information he was facing, and he was *already* unable to compute acceptable responses as is, even without me trying to dump more on his plate. The only thing left for me to do was to help him figure wtf was going on, and, if possible, what he might want to do about it if we could get all things considered. This is the

point at which we've exhausted our other options, and have to put in that work if we're going to do anything useful at all.

So I listened. And took him seriously. And I tried to figure out wtf was going on.

It wasn't easy. I never did figure it out. He threw me some curve balls that I really didn't know what to do with. When someone tells you that you look like the guy his girlfriend cheated on him with, and tells you not to look at him, wtf do you do with *that*? Avert eye contact? And lose the ability to look at his face to understand where that's coming from? Do you listen better by looking at him, or by not looking at him? Is it worth bidding for security *at least this far*?

Sometimes "listening" and offering attention requires listening and attending to what they want to give you, and the safest thing is to just accept that you won't know, and try to figure out from what they *will* give you. I think I sensed that in the moment, because that's what I did.

Although I didn't figure it out, I did buy enough time for his roommate to show up and cheer him up, and I don't see anything I could hope to do better to this day. It did work out though, and when I saw him again a year later he told me that he did really appreciate it, so I'm glad that even what felt kinda useless meant something to him. Effort isn't everything, but I guess it counts for something.¹²

The only other option, after escalating from attention through security and not finding suitable answer, is to not try. To just say "Well, this is more work than it's worth, at least right now", and disengage. Either you know enough about what is going on to navigate through security respect and attention to some sort of resolution with the amount of time and effort is available and it's worth, or you don't.

Are you willing to let what happens happen, if you don't step up and engage with the situation and all it brings you?

Notes

1. I'm drawing a distinction here between "irrationality" and "uncertainty". If you're failing to update on the evidence that you have and which you're physiologically capable of updating on, you're irrational.

If you're updating, but the updates look "stupid" and "irrational" to someone operating from a different set of beliefs, then you may be wrong and you may be ignorant but the problem is not a failure to do rationality. ↪

2. This makes for a good TAP. When you notice yourself doing insecurity over a "What if!?" try it on as a "Okay. What if?".

More generally but more challengingly, any time someone says something offensive like "I disagree" or "here's what you're missing" any displeasure at hearing this serves as a good question for asking "Okay. What potential fact of reality am I taking offense to right now" and, if you can handle it "Is it true?". This person thinks their disagreement *means* something. Okay, does it? They're acting like they're *above me!* Okay, *are they*, in the ways relevant to this statement?

If asking that question itself brings pang of displeasure because "What if!?" (or the equivalent on the other side of the coin "NO! THEY CAN'T THINK THAT!"), then we're back to "Okay, but what if?" ↪

3. Or again, the other part of ourselves. ↪

4. Getting your flu shot should feel like "not getting the flu", which a weird one to put into experiential terms, but it's not gonna be a *bad* experience. ↪

5. Discomfort at seeing someone uncomfortable is often carelessly conflated with "caring", but the object of the care is not the other person out there in reality. When the thing you're working to change is *their feeling*, you're not caring about their *actual* wellbeing -- and therefore you're not giving them reason to change their feelings either. Trying to "help them feel better" is often nothing but "trying to help *ourselves* feel better", by pushing away their feelings. This avoidance of the discomforting reality comes at the cost of those we purport to care about *and* their feelings.

I catch myself here not because "I'm supposed to do this other thing to be effective", but because *I care about her dammit*. When I noticed that my instinct was to "make the feeling go away", I noticed that I was flinching. When I notice the flinch, and attend to the question of whether I needed to hide from the feeling, the answer was "absolutely not" -- and that is the move that fixes the flinch. ↪

6. Which I'm pretty sure she actually did. ↪

7. I felt a bit weird about "confirming" what I absolutely 100% knew, but felt I should verify before stating as unqualified fact that she took it how I intended. ↪

8. And this is why I'm passionate about this stuff. It's not just "Hey look, this is an interesting and effective way of helping someone deal with air travel anxiety that isn't gonna hurt them anyway". It's that when you sacrifice truth and honesty out of incautious thinking, you have *no idea* what you might be giving up.

I don't know what would have happened if I had not had this conversation, and I don't think I would have been able to if I hadn't invited this woman to find her limits, and then backed off once we found them together. There's a good chance they'd have ended up together anyway, because they were just so clearly right for each other. But I'm not going to take that risk just so that I can avoid coming off like a jerk to someone who can't yet see the deliberate work, and risk, and *care* that goes into it. To the extent that the motivation is not obvious, it's just an opportunity to talk about a topic that is dear to me and why it is so.

Similarly, the air travel anxiety may have been a small issue in itself, but the ability to leverage security created by that kind of discomfort-facing very much *is not*. It's always difficult to tell without a control group, but the impression I get is that there were a couple instances where being able to lean on security in that relationship were at least as significant. ↪

9. That's because we're essentially doing what Frank Farrelly did in Provocative Therapy. If you're trying to help someone who isn't already asking for your advice because you're Professional Therapist, you will need to negotiate for respect while maintaining security. You will be cocky and bold, while empathetic and playing devils advocate. It's no coincidence that Provocative Therapy was developed when working with involuntarily committed patients.

In contexts where a person seeks out therapy from a therapist, the respect is already there or they wouldn't be seeking the therapists input. In such cases, if the client ever acts out of concordance with this, bringing thing

back in line just takes saying "Oh. I thought you were here because you wanted my input. What are you here for?" without passive aggression -- so there's generally little place for provocation. ↩

10. That bus cartoon is the worst. The day before my first daughter got that next flu shot she had been excited to get, she watched that stupid bus show that tries to teach kids about shots, and I had to start from scratch. The makers of that show teach their own perspective on needles, which is "BIG SCARY BAD BUT WE HAVE TO PRETEND OTHERWISE", so they showed big and scary needles with everyone afraid of them, and some adult coming in to gaslight them about how it's actually okay.

Surprise surprise, my daughter updated on their actual belief rather than falling for the lies that they advertised to be lies". (facepalm) ↩

11. As the underlying principles are *beneath psychology*, and therefore species agnostic, this process of bidding for security works great with dogs as well -- if you're willing to risk getting bit.

I do have some scars on my thumb from misjudging that once and playing too roughly too quickly, but most of the time dogs just relax into the play and love you for it. ↩

12. In cases where you don't have any of the answers, and what matters is how the other person conceptualizes the world, what comes out looks a whole lot like Rogerian/person centered therapy. ↩

Part: Clarity

Clarity grounds these bids in reality, and shows when bids for security are safe and worthwhile.

The Frustrations and Perils of Navigating Blind to Rocks

Over the course of [this sequence](#) I've been building up to the non-obvious idea that when you know¹ the truth you can basically just point at it and people will believe you -- even if they think they "can't" because "it's subconscious" or "it's irrational and won't listen to reason!" or whatever.

First, we showed how when people already respect you, you can signal "I'm serious, just trust me on this and follow my lead" [and get results that they wouldn't have thought possible](#), like getting burned without pain, constricting blood vessels to stop bleeding, etc. Say things like you [expect](#) to be followed, and when the prerequisite of respect is there, people will.

Then we showed how [respect is actually really simple to earn](#), provided that people don't flinch from engagement. Basically, don't flinch from engagement yourself, and don't change your mind. The examples I gave weren't any more complicated than "Let me know when you're ready" and simply waiting for a few seconds without additional input.

Finally, we showed that the security needed for people to engage with reality can also be built by simply engaging with reality oneself, while caring about them. Or appearing to, at least. We can only go by how things as they appear to us, so what if it's faked?

This raises the question for the post today:

If security can be built as simply as "Oh yeah, that's really scary" while smiling...

And if respect can be built as simply as *saying nothing* instead of rushing to say "It's okay, I don't respect me either!"

And if attention can then be directed by just telling people what will be true...

Then how is this not trivially abusable? Or is it? What's to stop someone from just putting on the body language and saying the words?

The superficial answer is that non-psychopaths have a "psychological block" against this kind of manipulation. It's superficial because the whole point of this sequence is that "psychological blocks" dissolve once you recognize them to be embodied beliefs about the world which respond to evidence. If you actually knew that faking things is in your best interest, you'd do it no problem. In other words, you almost certainly recognize that simply faking these things would be a poor bet for you, even if you don't also recognize what you're picking up on.

So what *are* you picking up on? What happens *if* you try to fake it? Why *doesn't* that lead to good outcomes -- for the person who does it at least?

The fundamental problem with faking security

Well, the thing you'd be faking *first*, is security. If you're sailing in dark murky waters with no knowledge about the location of submerged rocks, there's technically nothing stopping you from going full steam ahead with a smile on your face. The only potential problem with that is that you might sink. So long as you'd prefer strongly not to sink, you'll probably feel tentative navigating these unfamiliar waters - and that's a good thing.

So long as we can see the threats, or lack thereof, then we can steer around them and go full steam ahead. If someone is angry and threatening violence, we have a choice whether to keep our defenses up and threaten them in return. If we reflect and amplify these "violent neutrons", we might not de-escalate, but we'll at least be prepared for the battle.

Our other option is to lay down our spears and our shields, and open ourselves to cooperation by attending to the invitation to cooperate to the exclusion of the threat of conflict. This can be a powerful de-escalation technique, and it leaves us vulnerable if things turn confrontational. Not "but", "and". *That's how it works*. It works because it is a *credible* display that we don't expect or intend conflict.²

We get to act with security when we accept these risks, and we want to accept these risks when we can see that it is a good deal. When it's your toddler who is angry, it's easy because the rocks are pretty small and pretty soft; it's clear the kid can't hurt you. When it's a full grown man, whether it makes sense to act with security depends a lot on whether we know if the person is armed, how far they will escalate and in which specific cases, *whether we know how to handle ourselves should it get physical*, etc. If we know how to see a punch coming, and slip to the side before it gets there, getting punched at is less concern. Some people can see this so clearly that they keep their guard down *in professional boxing matches*. Seeing what it's like to *be* the other person allows security to pay off for a deeper reason. Instead of dancing around just out of range of *physical* punches, and slipping to the outside when we do provoke them, we get to *dance around the sensitive topics without provoking outbursts in the first place*.

Clarity enables security, which opens the path to cooperation and agreement over the direction of attention. The reason we can't just bid strongly out of stubbornness is because without clarity about what will actually happen, we don't know where we will end up. And that's a *problem*.

I'll illustrate with a few stories.

My friend told me of a man her friend group referred to as "The Wizard". He had such a way with words, and would captivate attention so *magically*, that he earned the nickname of "The Wizard" and people would always find themselves doing exactly as he intended³. On the surface, this sounds great! What's not to like about people doing exactly what you want? This man will never have to go through the heartbreak of a breakup, because he can persuade her to stay with him and be happy about it. Even if his girlfriend *did* break up with him, he could always persuade her to take him back. And who *wouldn't* want to avoid the pain of heartbreak?

Except it didn't work out that way for him. It was true that he would have been able to talk her into taking him back, and that's exactly why she broke up with him over E-mail. His impressive skill in bidding for attention without provoking respect or security issues brought him great results, until it didn't. And when it didn't, it cost him the opportunity to correct his mistakes and repair his relationships. He got the attention, but not the respect and security needed to believe he was *right* to get the attention.

This is the cost of bidding effectively without backing clarity. It works until it doesn't, and the question is over how far you get before it doesn't.

David Koresh, and cult leaders in general, show a similar pattern. They claim to speak for God which is about as high of a level as respect as can be claimed, and sometimes some people buy it. Then he gets to make bids for absurd things like their wives. Which, again, can be superficially appealing but remember how his story ends. At 33, in flames, along with his wives and children. If people around him *hadn't* granted him such absurd bids for request, or if he had been wise enough to not make them without better grounding, he might be alive today, telling his grandchildren about how he could have been the next Jim Jones -- and how he's glad that he chose not to be.

Donald Trump is an interesting and culturally important figure of today⁴. Back in 2016 when he was working his way up the Republican primaries it stood out to me that he was bidding *hard* in this respect game, provoking insecurity left and right and then using this to elevate himself. I didn't know if this would *work*, or if the defense mechanisms of the political system had a way of shutting it down, but I knew that it seemed to be working and that people didn't understand or appreciate what he was doing or why it was working. Scott Adams, another "trained hypnotist", spotted the same thing and got loud about it. Adams made the point that Trump was showing *skill*, and this is what he's *supposed* to do, and once he wins he will transition to a unifying stance and the whole country will love him and everyone will clap.

And then that didn't happen.

In my terms, Adams was predicting⁵ that Trump would transition from bidding on respect when he needed respect to bidding on security when he needed security. That would have been the

wise thing to do, and he's correct that provocation for respect to the detriment of security *was* what he needed to be doing earlier on and therefore can't actually be used as evidence of incompetence -- as evidenced by the fact that it *worked*.

The trick is that it's also consistent with another hypothesis, which is that Trump knows how to play the respect game well because at the end of the day, he knows which bluffs he can get away with -- but not what to do when he gets to the end of that road. Bidding for respect, without the clarity to back it up, can lead to situations where no matter what Trump walks away thinking he's great, and either other people believe it or they don't. Before 2016 some people close to him believed it, lots of people didn't take him that seriously, and he would just write them off like they would write him off. As president, we can't just "walk away from him", so we're pressured to resolve this, and so instead of ending with "agree to disagree" it provokes very real insecurities.

And for good reason! People really don't know how to orient to the things he's saying *without* accepting things that seem harmful and false to them, so *of course* they're not going to take him on face value. He could just be bluffing, and granting those bids legitimacy would be a mistake, in expectation. Without negotiating security, without *showing* people that his stances really do resolve to something that is knowably good, so that they can feel secure, they can't know that they do and as a result they can't do security and agree on the terms for honest engagement.

And neither can he, really. Because how can you know that you're on the right side if you don't understand the validity behind your oppositions objections? And if you do see it clearly, *why not show it*? No one thinks that Trump wouldn't love to be loved and respected by the *whole* country instead of just enough to win an election, when combined with those who are willing to hold their nose so as to not elect the opposition who seems even worse to them -- so it's in his best interest to do so. In absence of doing it, we can only conclude that he doesn't see how, and that leaves us as a country split on whether he's a force for good or a force for evil. But because he played the respect game effectively, both the [left](#) and the [right](#) agree that what Trump says is super important and the 2024 election determines the fate of our democracy.

Navigating without clarity about how things ground out can get you far. It can get you to the presidency, even. But it will be a controversial one, and if the things you're missing turn out to be important, maybe a horrible one. Maybe those things won't turn out to be important, or maybe they will. Isn't this exciting?⁶

Things are better in expectation if we don't disconnect ourselves from feedback to the point where can bid without bound. Maybe not in the myopic short term, but definitely in the longer term. And we tend to sense that, so we generally don't.

But that's the easy part. The importance of clarity in one's map of the rocks is as obvious as "Hitting rock bad. Must see rock".

The importance in clarity in our maps of our maps gets much more tricky. When we lack clarity about what our own maps contain, we end up completely flummoxed by simple problems like [getting out of cars](#) and [looking up from phones](#), thinking instead we've proved that looking up doesn't work, after all, we *empirically tested* [our own understanding], and found that it was incorrect.

This will be the topic of the next post.

Notes

1. Truly *know*, not just "Have a narrative that you're very very attached to, and [which proclaims itself to be obviously right and well evidenced](#)"!

This is a much much higher bar than we usually hold ourselves to. ↩

2. If you're afraid of a dog that's growling at you, the *last* thing you probably do is lie down on your back and offer it the back of your outstretched hand for it to come up and put its mouth to -- because that gives it a chance to bite your hand when you're in a position where you can't do *anything* to fight back. Which is why it's such an effective way of getting dogs comfortable with you. ↩
3. He's a natural hypnotist, essentially. ↩
4. Let's see if I can side step the laser beams myself, saying only things that both sides will agree with yet are also illuminating. ↩
5. I don't buy for a second that Adams actually believed it, just that he was overextending his own bets for whatever reasons. ↩
6. I understand that most people do not see it as a fun exciting thing, but as I write it I do. It is an invitation to come have fun, and tune in to reality to find out which is right. Obviously it's better to figure these out at lower stakes, and perhaps we could have sorted out our security issues *before* the elections if we had clear thinking on what insecurity *means*. And we didn't, so this is the reality that exists for us to play with. ↩

How to actually decide

Normally, when we think about being less wrong, we think about being less wrong about external reality. When sailing, we ask questions like "Where are we?", "Where is our destination?" and "[Where are the rocks hiding beneath the surface?](#)", and really hope we don't get these things wrong. Because if the map by which we're navigating doesn't correspond to the territory, we're likely to end up very confused about what we're seeing, and lack the ability to navigate effectively.

But there's another kind of confusion which can be extremely tricky to navigate.

The chronic pain transcript is a good example of this other kind of confusion and how extremely tricky it can be. [Part 1](#) introduces the problem as essentially debilitating chronic pain from physical nerve damage, which sure doesn't *look* like a problem of confusion. Yet the solution, shown in [part 2](#) in full, reveals that careful navigation was able to substantially relieve his suffering overnight.

This "other kind" of confusion which can be so deceptively tricky, is confusion about the content of our own maps. This person suffering from chronic pain wasn't wrong about having nerve damage, but he misunderstood the cause of his suffering, what his pain meant to him, and why he wasn't attending to it. It's as if he was sailing towards the point that his map had labeled as the destination, but because he misunderstood his ship's motion in that direction as "due to the wind", he wasn't able to effectively get his ship to do what he thought was right.

When we mis-map our own maps, we end up equally unequipped to navigate, only instead of being confused about the reality that surrounds us, we will be confused about how we're trying to navigate the world around us. And instead of being ineffective at interacting with the outside world, we will be ineffective at interacting with our own beliefs, and methods of navigation. Which is how he ended up stuck suffering for so long. And how many of us get similarly stuck, in all sorts of circumstances.

When you're confused about the fact of whether you're confused, and hold the mistaken belief that you aren't mistaken about what your beliefs are, it can be a really difficult to escape from this epistemic trap. Beliefs like "pain is bad" and "I know what I believe" feel like directly experiencing truth, so it's really hard to find the seams to begin tugging at -- or to even know to look.

Rather than continuing to tear at the seams, let's begin building an alternative model, which does not have these limitations.

Can we "just decide"? What would it look like if we could?

Imagine that we have an irrational fear of dogs -- only, we're too damn naive to realize that you can't reason with irrational fears.

Obviously this is crazy, and it's not so easy. One time I was so scared that I literally couldn't get my hands to let go of the rope, and I had to use my feet to kick my own hands off. You've probably had your own experiences trying to "reason" with things like fear, so I don't think I need to sell you on the idea that "maybe just reason better" is naive. But let's explore it anyway for a moment, and just assume I was being dumb or something when I was kicking at my hands instead of just letting go.

How might this go? Often, people will make mistakes that lead to failure even when the task they're attempting *isn't* impossible, so let's first make sure not to do that. Assuming we're *extremely* epistemically careful, what is the first sign that we're butting heads with reality?

In order to answer this question, we must avoid a bunch of common pitfalls. For one, we must make sure we know which question we're asking. "Should I be afraid of this?" is not the same question as "Is it dangerous?". One implies the other, so it's extremely easy to conflate them either without noticing or without noticing that it matters, but they are still distinct questions. One is "Is this thing dangerous?", and the other is "Is that answer we just gave correct?".

Answer the former, and you answer the former. Okay, maybe our answer there wasn't correct. In order to correct our answer to the latter question though, we have to change our answer to the latter question.

Is this thing dangerous?

How it feels from the inside, and the temptation to flinch

Once we ask the right question, we still have to answer it if we're to change our minds. If we flinch away before we answer it, then our answer won't change. That's why it is important to understand [how to diagnose and correct a flinch](#).

This can be a very uncomfortable question. Put yourself in the shoes of someone struggling with an irrational fear, and notice what happens when you're asked "Is it dangerous?"

The responses that come up tend to be things like "No, but I can't help it", or "It's irrational, I know". The former answers and then immediately negates the meaningfulness of the answer, while the latter doesn't answer the question at all. Neither of them have the *gravity* of just saying -- sigh... "No. It is not dangerous", or "Yes. It is dangerous".

Or even "I don't know."

If we're going to be epistemically careful, we need to notice these flinches before we commit them, and refocus our intention on finding truth.

Is it dangerous?

The kernel of truth in the [perceived] error

If you're sufficiently afraid of dogs, and not happy to just answer "Yes!", this can be a very difficult question to answer so definitively like that. Because who wants to be afraid of dogs? Self judgement as "irrational" starts to make a lot of sense, sometimes. Often the answers we reach -- and which are shown by our fears -- *do not look correct*. For one, being afraid of dogs is no fun, so that doesn't seem like the right answer. But for two, you also know that it's pretty silly, since most dogs are completely sweet and harmless (most of the time).

At the same time, as overblown as the fear may (or may not) be, dogs bite. I've been bit even by "nice" doggies, and have scars to prove it. If you're phobic, maybe you've been bit too. If you give yourself some room to not worry about looking stupid and look at the facts, there's a reason to be afraid of dogs. You can't guarantee you won't get bit (again), and getting bit really freaks you out. You really don't want it to happen (again). Once you're honest about this you can start to frame it as a decision: you can avoid ever getting near dogs and live in fear of them... or you can go pet the dog and risk getting bit.

But just because you see it as a decision doesn't mean it's easy [or that you're ready](#) to face it. It's easy to get caught up in these kinds of things. Imagine someone with this fear comes up to you pleading for your help, and you're stuck in this "naïve" perspective where people can just do whatever they want, and just have to figure out what that is, exactly.

"I shouldn't be afraid of dogs, this fear is stupid and irrational".

Okay? So go pet the dog.

"I can't!"

Why not?

"Because I'm afraid!"

Afraid of what?

"It might bite me!"

Yeah, it might. Is that okay?

"Of course not! I don't want to get bit!"

Then don't pet the dog

"But I want to pet the dog!"

Okay, then go do it

"But I'm afraid!"

We can go on all day like this, bouncing back and forth between "I don't want to be afraid, that's stupid" and "Yo, I'd really rather not get bit here".

The NLP technique of "collapsing anchors"

NLP has a neat technique to deal with this kind of thing called "collapsing anchors", which uses another technique called "anchoring". "Anchoring" is more or less just NLP's term for "classical conditioning"¹, however their approach is notably different in that instead of merely using it to explain behavior they use it in a very hands-on practical sense to actually change behavior.

A more thorough description can be found with a quick Google, but the brief summary is this:

- 1) Fire up state A and "anchor" it to stimulus 1 by presenting the stimulus at the same time. E.g. "Think of a time when you felt confident. As you step back into that situation, see what you see, hear what you hear, feel what you feel", and then once they're in that state, anchor it by touching their right shoulder.
- 2) Break state. Ask an unrelated question or whatever. How are you liking this sequence so far? Shoot, I was hoping it'd make more sense than that.
- 3) Fire up state B and anchor it to stimulus 2 in the same way as before. E.g. anxiety in the problem context, and their left shoulder. Maybe test again
- 4) Break state.
- 5) Touch both shoulders simultaneously, and watch what happens.

It actually is pretty neat, and you might want to play with it in that form first, just to get a feel for it. Especially if you take a moment to test your anchors to verify that you got that part down before trying to collapse them into each other.

The way it's usually explained is like you fill up one bucket with "confidence" and the other with the problematic anxiety, but you make sure to fill the "confidence" bucket higher so that when you mix them together it overwhelms and drowns out the anxiety. One simple trick for confidence in any scenario!

Except of course it doesn't actually work like that. I mean, it does *work*, sometimes, but when it works that's not really why.

Properly understood, "collapsing anchors" is a set of handrails intended to guide people along the path of *actually deciding*.

Imagine you have a speech to give at your friends wedding or whatever. You're feeling nervous about it, and you don't like it one bit. Then someone asks you "What happens when you try being confident?". What's your response? "I dunno"? "I can't"? "I'm not confident"? "Screw you"? Probably not anything too profound or helpful.

But what if things get spelled out in a bit more detail?

When was the last time you did something noteworthy, something potentially challenging and intimidating, but did so *without* fear or anxiety? Either because you had practiced so many times and succeeded so many times that there's just not that much uncertainty left, or else because there's just nothing left to do with whatever uncertainty is left. Something where you don't "feel confident", you feel *competent* -- because when you think "Am I actually good at this, or am I just propping myself up?" you immediately know "No, I'm actually good at this" -- in the same way that you'd answer "Are you *sure* you have a nose, or might you be misremembering" with "I have a nose. It's right there, I can see it". You know what it feels like to just *know* that you're ready to handle whatever is about to come at you? That's an experience you're familiar with, yeah?²

Now, hypothetically, what would it be like if you were somehow able to be in *that* state of mind with regards to this speech you have to give. What if you had that same *confidence*. That same *knowing what you're doing*. That same certainty that things will work out because you're ready to *make* them work out. Only this time, in the context that used to flummox you. How does it play out differently this time? What might happen, if you were to simply trust in yourself and just go for it?

That's a bit different than simply asking "Have you tried being confident?". Now, the answer isn't necessarily "Confidence bucket bigger! Confidence bucket win!". Maybe it's "Lol, I'd probably be careless and mess up again like I did last time. That's how I lost my confidence in the first place" -- but even that is an actual answer.

Or maybe it's "Huh... I've never thought of it that way before", and you have something new to take to the problem context, which may or may not be a full or partial solution.

The important part of "collapsing anchors" is that you're making sure that you're actually summoning the embodied understanding that leads to these disparate conclusions, and pulling them into the same space so that they can be considered *simultaneously*. Rather than simply

"saying words" and wiping our hands of it as if "I did *my* job", actually making sure the words -- and gestures, and entire context -- actually summon the experience you're trying to draw from, so that questions like "What if you tried being confident?" get answered from the first person, as you *find out* what would happen according to your object level predictions. As opposed to subtly switching questions from "What would happen?" to "which description of what would happen would be true?" -- or dismissing the question entirely before even trying to answer it.

When done right, "collapsing anchors" brings things together, in a way that "Have you tried just being confident" generally doesn't. And in order for the two perspectives to be merged and chosen between, *someone* has to do the legwork to actually connect the handles like "confidence" to the felt senses to which they refer.

NLP's technique of using kinesthetic ("touch") anchor is simply one way of doing it³ -- one simplified set of hand rails to introduce us to the concept of "actually deciding" in a simplified way so that we can follow. The kinesthetic anchors aren't really necessary. Above I used italicizing and multiple phrases referencing the state I was aiming to evoke to try to fill the same role. Maybe it worked for you, or maybe it didn't really connect -- it's hard to "force" people into engagement in this format, and it does help to have touch and body language and real time calibration -- all of which are useful. Rather than thinking of it as a robotic "touch same spot to elicit same feeling", recognize that it's a form of *language*. It's communication, and by leaning on and emphasizing certain key words (like "that *confidence*"), you convey that you're linking that word -- and/or that look, or that touch -- to that state and using it as a handle. You're digging in until you actually have a hold of the entire experience that goes along with it, and pulling the two seemingly mutually exclusive experiences/conceptions of reality into the same room and saying "Hey you two, what do you think of each other?". Because our experience of the world is our map of the world. It's what we use to navigate because it's what we perceive as *real*. And the verbal strings we call "beliefs" are often meta-maps, declaring what our maps *should* say, or which beliefs would be *true*. And if we want to change how we navigate we have to change our object level perceptions of reality.

Sometimes it's enough to say "and what would that be like if instead of feeling anxious you had that same confidence that you take to <thing you're good at>". Sometimes you have to spend more effort reminding people what it even feels like. Every once in a while you'll even come across someone who is sufficiently on top of their shit that you can just say "Bro, have you even *considered* being confident?"⁴ and they'll do all the legwork themselves.

And of course, sometimes people flat out don't wanna do it for one reason or another. But let's look at what happens when you do, as applied to our old fear of doggies.

Deciding whether dogs are too dangerous

You don't want to waste your life living in fear of nice doggies, yet you also want to not get bit. You really want to have them *both*, and wouldn't that be nice. It'd really be great if you didn't have to "make up your mind" about which is worse: not getting to pet doggies and feeling afraid all the time, or taking a risk and maybe getting bit. And so you don't weigh the two felt senses⁵ and collapse them together. Getting the best of both worlds is always good, and it's a good idea to think through whether maybe we *can* have our cake and eat it too -- at least to an extent. Maybe we can discriminate between the smiling Labrador puppy and your friend's snarling rescue dog that looks suspiciously like a wolf. But at the end of the day labs can bite too, and at some point you need to weigh your options and decide.

And when you do, it goes something like this...

I want to pet the doggie, and if I do, I might get bit.

...

...

...

(Seriously, give it a moment. Shit takes time.)

...

Is it worth it?

Am I willing to stick my hand out and pet that dog knowing that there is some chance that the dog is going to bite it?

And then you sigh a bit. And then you're silent. And you picture not the separate issues of petting (good!) and being bit (bad!) but the combination package of getting to pet the dog and maybe but probably not (but maybe) getting bit. It doesn't yet *have* a goodness value, since you have not assigned one yet. You've never weighed the ingredients and settled on a value for the package deal. You've never decided.

So it takes some sitting. And some silence. And some passively imagining the package deal and asking yourself "do I want that?". It's a very gut level thing. You can watch it happen. You can notice that not all of the factors of emotional weight are on the table and go fetch them. But it's not your job to force it, or to try to tell yourself what you want -- if you tried that with anyone else people would call you a controlling jerk and they'd be right. Proclaiming "[I declare bankruptcy!](#)" is not the action of declaring bankruptcy, and proclaiming "I have decided!" is not

the action of deciding. No, we're just asking ourselves -- and letting ourselves decide whatever feels honest "Is that actually something I want?".

And when the answer comes to you there's no "but"s. It's not "I want to pet the doggy but I don't want to get bit", because it's not two things. We're not asking "Wouldn't you like to pet the doggy [if it could be known safe]?", and we recognize that to be a different question. An *easier* question, with an obvious answer. This time we're asking if you want the package deal of petting lots of doggies and *maybe* getting bit one in a thousand or so. We're asking about what you *don't* know. And when you have an answer to that, it's just an answer - "Yes, I want that deal" or "No, I do not want that deal".

If your answer is yes, then you can say "Yes, I want to pet the dog, even knowing that I might get bit. I still want to pet the dog because it's worth it. I want that package deal where my hand might get bit". You *want* your hand to get bit, probabilistically, because it's a small probability and overwhelmed by the positive puppy petting that it comes along with.⁶

Or if your answer is no, then you say "No, I don't want to pet the dog. It's not worth the chance of getting bit". And that's the end of it. It's not "But I wish I could pet it and it wouldn't bite me!". It *would* be nice if you could have that, yet you don't think to say it because it's just "*duh*" - it's a dog and you can't predict them perfectly, duh. That "but I wish..." has no steam left after agreeing with it. Like "of course I wish! what's your point?". Like "Wouldn't you like a million dollars to just fall out of the sky?" Sure? who wouldn't?

Yet the thought doesn't hold your attention.

And either way, there's no conflict. No two separate desires. Just a congruent choice coming from a decision you had not made before.⁷

At least, that's what it'd look like if this naive approach to solving irrational fears actually worked. In the next post, we'll go over what actually happened when my friend came to me asking for help with her irrational fear of heights.

Notes

1. E.g. Pavlov "anchored" anticipation of dinner and the associated responses to the sound of a bell. ↩
2. Notice how [this only works if it's coming from someone you respect](#)? If you don't, this sounds like silly shit you probably don't intend to take very seriously.

"You sound like some wannabe NLP kook!" ↩

3. NLP's "collapsing anchors" falls out when people are secure enough to face a decision but need explicit help noticing and bringing together the relevant experiences that inform that decision, because they didn't realize there was something they weren't doing. ↩
4. Or "Apparently you can just decide to not swell injuries". ↩
5. In the Gendlin sense. ↩
6. I wish I had a reference to cite here, because a lot of this is taken fairly directly from Joe Fobes's "What Do You Want?" protocol. Essentially, he noticed that by refusing to validate the desire to eat ones cake and have it too as a reason to get stuck, you can actually resolve such issues quite effectively.

It was his insight that the issue is fundamentally that of conflicting desires and an undecided mind, that the resolution involves seeing the realistic option as *one thing*, and that this entails *wanting* the probabilistic bad outcome too. I think he even used the example of a dog bite when teaching it to me. His methods were quite crude at the time, basically pretending to not understand and just annoyingly repeating WDYW? until people would get fatigued of fighting for what can't have and make up their mind, but it was still surprisingly effective and was a great proof of concept for this key insight.

Where he went next with it is a direct improvement as it doesn't require pretending to not see the desire for both, and as a result is more effective. I've gotten a lot of mileage out of this protocol when I have gotten stuck, and it is exactly this protocol, applied to pain, which led me to see first hand how pain *actually* is just information -- that *matters* to us. Rereading now, there's clear bits of this acknowledgement routine shining through too, actually. ↩
7. The insight that a lot of apparently irrational behavior actually has fairly valid reasons of which we are unaware is not unique to me, and is a recurring theme in lots of therapeutic schools of thought. ↩

Solving irrational fear as deciding: A worked example

[Last post](#) I sketched some of the mechanics of how we actually make decisions and change our mind, as well as some ways we fail to complete this process and what it would feel like to get back on the track to make a decision we've been failing to make. In this post, I'll give an example of when I walked a friend through this process.

A friend of mine once told me that she had a fear of heights that would come up when rock climbing. It wasn't extreme, and she would generally just push through it, but she couldn't shake it. "It's irrational", she said. She'd had it for years, and made no progress reasoning with it -- the CBT she was so fond of and had so much success with wasn't working here. She wanted to learn how to deal with irrational fears (especially so that she could help others), so I told her to come visit the rock wall I had made with a friend, and we'd see what we could do.

Our rock wall had the advantage of not being at a real climbing gym. It was on the wall of my friends workshop, with a bare concrete floor. There was nothing screaming "This isn't safe to climb", but perhaps not all of the usual signs that it is either. Once she made it to the top of the wall, I started having some fun with her fear. "Oh, by the way, can you give that concrete anchor a shake to make sure it's solid? I want to make sure it's not coming loose again.". "Good thing this fear is irrational, because that concrete floor is hard. Can you imagine what would happen if your head hit this concrete? I don't think you'd bounce... I'm sure you're fine though. Being afraid would be irrational".

How do you think she responded? How would *you* respond, in a situation like that?

The first thing that comes to mind, is "Screw you, I can tell you're trying to provoke me!". But keep in mind that *she asked for my help*. The reason she's up there in the first place is that she wanted to get over her irrational fear, and to learn how it works. That's why she was up there, and that's why I was teasing her. If she were to change her mind on that, that's fine with me. She knew my intent wasn't jerkish, which makes it hard to claim jerkish intent, so "Screw you" doesn't really fit anymore. And without that, what's left? "I can tell you're trying to provoke me?". Yeah, congrats for noticing ;). She was too smart to walk into that one.¹

Another possibility that comes to mind is "Get me down!". But again, she's up there for a reason. She can ask for me to let her down, and I would, and I would congratulate her on her newfound realization that her fear isn't irrational after all -- because after all, if the danger isn't *real*, then why are you *acting like it is*? She could see that one coming too, and wasn't ready to accept that the fear was rational and worth cowing to.

What other option *is there*, besides staying up there and letting me tease her, which doesn't have similar problems?

She stayed up there and let me tease her, fully grasping the point. If her fear is irrational, I shouldn't be able to get to her. She's free to take it seriously any time she wants, but I'm going to *notice* when she does, she's going to notice me noticing, and she's gonna have to find a way to square that.

The challenge, in cases like this, is in mind-reading what the fear is really about. If she's afraid that the concrete anchors will fail, and I incorrectly assume that her fear is about my potential inattentiveness during belaying, then all I could do is say "Well good thing I'm paying attention, right?" and she'd be able to say "Yes, but it's still scary" with no conflict. Because I wouldn't have addressed the actual cause of the fear.

It's a bit like playing battleship. Poking and prodding semi-blindly and empirically *finding out* what she believes which is leading to fear. We're poking and prodding to see what connects, as those are her *revealed beliefs* about what is dangerous.

If I were to ask her something stupid like... "What if a meteor comes down and blasts through the rope, causing you to fall!?", she could just say "I think I'll take my chances", and roll her eyes at the idea that this is a real threat. When I asked if she was *sure* her knot was tied correctly, she didn't find it so funny. That question connected with the fear, and she could sense it. She can either say "No", thereby admitting that maybe her fear wasn't entirely irrational, or say "Yes" and risk getting caught in a fib, or try to distract and deflect.

The conversation went something like this:

"You did check my knot, right? Are you *sure* that it's tied correctly"

"You wouldn't have put me up here if it weren't tied right"²

"Yeah, probably not, you're right. It's not like I ever make stupid mistakes... right?"³

Every time she'd deflect I'd just redirect her to the questions of importance: "How do you actually know you're safe up there? *Or do you?*"

How does she square her idea that her fear is "irrational", with the fact that I am *obviously* error prone and that she didn't even check my knot?

She could always say "Shoot. Yeah, I screwed up didn't I" -- if she's ready to admit there might be something to her fear.

Hypothetically, she could say "You obviously make your fair share of dumb mistakes, but you're also smart enough to know when you can't get away with them, so I know you triple checked my knot" -- if it really was that clear to her that it didn't call for checking my work. Or thinking about it explicitly, even. But that means accepting the risk, because if she says that and dares look down to check the knot, she's getting called out on it and she knows it.

The other thing is that unless you've sat down and thought these things through ahead of time, it's kinda hard to come up with explanations that make that much sense -- and when you do, you often find them convincing. If you don't, then it just kicks the can one step further down the road ("Yeah, probably.... How do you feel about betting your skull integrity on that?").

We went through this process for everything I could think of. Are you sure the knot is tied right? Are you sure this guy you don't know set the concrete anchors right? Are you sure the rope isn't going to fail because it's been stored in a bucket of harsh chemicals or something? Did you put the harness on right?

Every time I caught a glimpse of a fear response -- every time she couldn't *convincingly* act like it wasn't an issue, I'd dig in and poke around until we figured out what's going on. Okay, sure, I'm pretty aware of when I can and can't get away with making stupid mistakes, but she really should have checked the knot I tied. Fair enough. Okay, yeah, Jimmy knows whether his friend knows how to set concrete anchors, they've tested them, and there's more than one for redundancy. Over and over, we went through figuring out "Is *this particular fear component* rational, in part, or is it completely un concerning?"

The knot thing concerned her at first, but after thinking it through, it's just "No, *of course* he's not going to be making fun of me for not checking the knot without triple checking himself. The knot is safe". The anchors, rope, etc, all had some slight component to the fear, at first. After looking into them, and realizing that *in this case* there's no danger, it just kinda felt like... there was no danger. At least, with respect to *these* things failing.⁴

"Okay, so it can't be the knot, or the anchors, or the rope, or the harness. But clearly there must be *something* we're missing. There must be *something* that can go wrong here. It can't just be *safe*... right!?"

Same thing, generalized. She's free to feel fear and I'll commend her on her newfound wisdom and caution. Or she can realize that "Okay, fine, no there isn't anything else that could reasonably happen".

"But what if there's an earth quake! The wall could fall on you!". Sure, but at that point the universe is just out to get me, so might as well have fun while it lasts.

She can't *possibly* win, because it's heads I win, tails she loses. Either she takes her fear seriously -- in which case I win. Or she doesn't -- in which case... I still win. There's no such thing as a thing that is dangerous and also not dangerous *by the same judgement* -- and the judgement was hers. This is what I mean when I say that what I'm aiming to convey is [nearly tautological](#). There was no chance of "failing" here because I wasn't trying to get her to not be afraid. Despite her request to help her do so, I refused to write the bottom line first.

Instead, I rigged the game by making sure not to sign up to do anything I might not be able to do, and only sat back to point out where she *had to* be wrong. She had to be wrong because she was saying "I have an irrational fear", and that can't *possibly* be justified from her perspective. What she's saying what amounts to "Heights are dangerous, but my belief that heights are dangerous is wrong".

These [Moore's paradox](#) sentences are obviously absurd when stated plainly like that, to the point where it seems paradoxical. Yet people do this *all the time*. We just like to reword things a bit to launder our irrationality. "It's not *me* that's irrational, it's my *subconscious*. Yeah, that's it! I'm not the crazy one, it's *that other entity in my head!*". So long as you see it as "I have an irrational fear generated by mysterious subconscious processes over which I have no control" + "I know heights aren't dangerous", there's no contradiction to resolve.

But once we're secure enough to risk offense and point and look at what the revealed beliefs actually show, things change. All of a sudden, we recognize that she's *acting* like her beliefs are "Heights are dangerous", "That belief is wrong", "I don't have that belief". No "subconscious", no "*I* believe one thing but *my fear* believes something else" -- just a mismatch between object and meta level beliefs, and a mismapping of her own beliefs which steers her away from noticing that she has something to resolve.

And once you notice that you're just disagreeing with yourself, you start wondering...

How do you actually know you're safe up there? *Or do you?*⁵

What do you think happened to her fear, by the time she came back down?

What can *possibly* happen to her fear, provided that she's too smart and secure to miss the point?

Do you think she said "Hm, you're right, these are completely reasonable concerns to have... and it's still irrational"?

Do you think she laughed at *every conceivable danger*, including *the danger of the unseen...* and was *scared*?

Or do you think maybe she started taking her fears a little more seriously, thinking them through a little more, and making her mind up about which risks were worth taking, in each specific circumstance?

It shouldn't come as too big a surprise, but it was the latter.

Each different context has different factors to consider, and just because you don't need to be scared in one situation doesn't mean you don't have to be scared in another. It took a month or so for her to figure out the answer in her various climbing contexts, and by the end of a month her climbing on the wall was fluid in a way she hadn't realized she was missing.

She *thought* she knew how much caution she needed, but she disagreed.

It was never that her fear was unreasonable, just that she had never set aside the bottom line and actually reasoned about it.

Notes

1. Notice how this is only possible with significant [security](#)? If she were to freak out and fail to remember why she's up there, I'd have had to slow down considerably to ensure that she could keep all the relevant contradictions in mind at the same time.

The prerequisites of security and respect still need to be earned before you can direct attention like this. The point of this post, is to show an example of how once you have security and respect, you can systematically direct attention towards reality enable people to make decisions they were convinced they couldn't make. ↪

2. Notice how this doesn't actually answer the question? That's because it's a flinch. Saying either "Yes, I am sure" or "No, I am not sure" is uncomfortable, so she reached for an excuse to not answer instead -- and succeeded in delaying her decision.

It pays to be suspicious of your own motivations, when you find yourself not directly answering questions. If it really is a dumb question, we can always say "No, I'm not sure, and here's why it doesn't matter". Not "but", "and". ↪

3. Here I'm doing the opposite. I'm deliberately *not* flinching from her valid point that I wouldn't have put her up there without making sure the knot was tied correctly. It's true, so no harm in engaging with it -- if it throws me off then I *should* be thrown off, and take a moment to reorient to what I've just had a chance to learn.

The next step though, is right back to the thing she's dodging. Her own fear shows that she finds that to be not convincing, so maybe we can look at that *too*. And when you turn towards the possibility that things might not be perfectly safe, it's often easy to see why. ↪

4. At a glance, this looks similar to CBT in that we're looking at evidence showing the fear to be unfounded, and she's getting less afraid as a result. However, the *first* thing we do is almost the opposite of CBT.

I'm not not looking for evidence that it's safe, I'm looking for evidence that it's dangerous. I'm not trying to help her feel safe, I'm trying to help her feel afraid. I'm not trying to suggest that her fears might be irrational, but that they might be *rational*.

This is because the conditions that call for CBT didn't apply. If you have someone who hasn't already considered the evidence against their fears, and who was open to doing so, then CBT is a novel insight and falls right out of the framework. She already considered it and concluded that her fear was irrational, and was closed to considering the evidence that the fear was wrong because she thought she already did that. So the opposite of CBT is what fell out. ↩

5. It's worth pointing out that I was primarily *not* saying "Look, this is contradictory. You need to pick one", because convincing her of that would only convince her that she was doing it wrong -- which she kinda already knew. My focus was on "*is this actually dangerous or not?*", because that's the object level disagreement that she needed help resolving, and if she knew how to resolve it from meta observations alone she would not have needed help.

The meta level stuff came *after*, in the explanation of "Okay, so why did that work? What was I doing, exactly?" -- and a little bit of real time teasing between the lines like "You know that there's no way for you to get out of this without looking a little silly, right? :P"/"Are you starting to notice how I'm *systematically* shutting down your deflections and redirecting you to what you're avoiding?".

The object level teaches a new response. The meta level teaches how to teach oneself the new response, and is much easier done following an example. ↩

Part: Tying It Together

Then we tie it all together, and highlight the overarching theme.

Putting It All Together: A Concrete Guide to Navigating Disagreements, and Reconnecting With Reality

This is the capstone post for my sequence [Beneath Psychology](#). In this post, in addition to compressing down to the actionable takeaways, I will be attempting to close all the loops, connect all the dots, and cross the final inferential step that needed all those words to justify and elucidate.¹

As a result, this post is dense, and will likely take some time to digest. If you haven't been following along, it's unlikely that what I'm saying will make sense. Feel free to comment even if you haven't, and it doesn't, and I'll do my best to clear up any confusions as they come up.

Concrete guide for navigating disagreements

Now that we've looked at all the individual pieces, we can start putting them together into a concise and memorable framework to serve as a concrete guide to action. These are not "rules that must be followed", or "techniques to try". These are *coordinate axes*. If you want more respect so that people will listen when you talk, here's what points in that direction. If a certain direction doesn't feel appropriate, here's what that reveals about your object level beliefs. That kind of thing.

In other words, these are your *options* -- and descriptions of what you will *want* to do, made easier to notice.

There are three layers: Attention, Respect, Security. We negotiate all three layers simultaneously, constantly updating our beliefs about where on each spectrum we want to play, and bidding/offering/accepting bids accordingly.

Attention

Bidding strongly on attention requires pointing in one direction to the exclusion of others. If you fear that you might be wrong, you will communicate this caveat by splitting your attention between multiple options, and as a result you will split their attention as well. To see if you're ready to bid strongly for attention turn to the following question: "What if I'm wrong?". If you bid honestly and are wrong you will be *surprised*. Are you ready to risk this surprise? Would it even be surprising? Because if not then you know you weren't bidding honestly.²

Bidding strongly might provoke a break of rapport if they're not ready to consider bids this strong from you. If you want to keep rapport because you're not prepared to bid more strongly on the other levels, and you still think you know the answer, you can see if they'll agree. By offering them attention and asking if you've given enough, you give them a chance to notice that you have, and offer their attention in return. This other key question is "Am I missing anything?"³

Sometimes they'll bid in ways that don't seem right to you, and instead of escalating you can ask for reconsideration. This means flipping to the other side of the coin on both of these key questions: "Are you sure you're not missing anything? What if you're wrong?". The first is the meta level check "Do you trust your own judgement here", and the second is reconsidering the judgment itself. What if you're wrong? Is that something you're willing to risk? The former has the function of opening the latter up for reconsideration. If I've bid then I've already taken the stance that I'm willing to be wrong, and will be likely to simply assert "I'm not wrong". I *might not* have thought this through, and if you ask me how I'm so sure... I might realize that I want to reconsider -- *if* you direct me to the object level question again. Then again, if it's already obvious that I might be missing something, then you can skip this step and return straight to the object level question of what if I'm wrong.⁴

When you notice yourself not confident in bidding further -- because you recognize that *you might be wrong* -- you might find yourself getting *curious*. When you're curious, that's your cue to follow that curiosity, and *listen*. You will likely want to offer your attention, attune to the experience they're aiming to convey, and find out what they have to show.

Respect

Bidding strongly on respect requires that you believe your bids for attention will turn out justified or that theirs will not. This means that you can put yourself in the position to stick your neck out and risk humiliation, and feel confident taking the bet because you've already humbled yourself appropriately -- in expectation. Sure, your bet might be wrong on the margin. Are you ready for that? The key question that enables strong bidding for respect is "Am I ready to be humbled?". Are you able to face laughter with a smile on your face, because you understand where it's coming from, and that it doesn't impact the credibility of what you're saying? When you make a bid for respect and turn out to be wrong, you will be put back in your place. Will this feel humbling or humiliating? If the former, that's fine, you accepted the known risks and made your best bet in honest expectation. If the latter, you made a claim without accepting the risks, showing that your words weren't honest and your bet not optimal in your own object level judgement.⁵

Of course, bidding strongly on respect risks breaking rapport and escalating to divergence on the layer of security. Maybe you're not ready to deal with that. If you think you deserve more relative respect, and don't want to break rapport, again you can offer them the choice with room to think and evidence that you're right. If you offer them the respect they think they deserve, and don't demand more than they're willing to give you, they might realize their error on their own -- if they are indeed in error. The key question that enables this reflection is: "Am I taking your perspective seriously enough?"⁶

Sometimes they'll bid in ways that don't seem right to you, and instead of escalating you can ask for reconsideration by flipping the questions again: "Are you sure you're taking my perspective seriously enough? Are you ready to be humbled?". Of course, the connotations don't come across right on that latter question as stated, so realistically it's something more like "Are you aware of how you're coming across right now?".

If you don't think you have the most solid footing, you can offer respect instead. Look up to people who have earned it. Listen when they talk -- at least when they bid for attention. If you respect them, you will be willing to entertain uncomfortable things which conflict with your worldview *because you think it's important enough to be uncomfortable for*.

Security

Bidding strongly on security requires you to face risk of harm. You might hurt someone, by allowing them to mislead themselves into harm -- and you will lose trust if people can see you flinch away from accounting for this. You might start a fight, and they might hurt *you*. In order to bid strongly on security, you have to put yourself in their shoes and expect *safety*. You have to lower your shield, and your spear, and make yourself *vulnerable*. If you can't ground this in strength and clarity, the risk is real, and bluffing will come back to bite you sooner or later. The fundamental question relating to bidding on security is "Am I ready to make myself vulnerable and face the potential for harm without bracing?" -- in other words "Can I actually fully engage while completely enjoying every aspect of it without subtly flinching from something in a way they could conceivably call me on and make me uncomfortable?". If you bid honestly and are wrong, someone will get hurt. You will owe someone an apology. Are you ready to grieve that loss and offer that apology? Will you actually be sorry? Because if not then you know you weren't bidding honestly.⁷

Of course, doing things that frighten people might cause actual harm. If you still think more security is likely appropriate and don't want to risk someone getting hurt if you're wrong, once again you can offer them the decision and evidence that things might be safer than they realize. The relevant question here is "Am I being too threatening here?" -- or, more naturally phrased: "Are you okay?"⁸

When people bid that things aren't okay and you don't want to just bid stronger, you can again flip the questions on them. Are you being too threatening here? Am *I* okay, in your eyes? Or, perhaps more naturally "Will you forgive me?".⁹

Maybe you know they aren't, and still want to help. If this is the case, you have the option to walk on egg shells and make really sure that nothing you do is threatening. You don't make fun of people, you don't expect them to listen to you. You don't expect respect, you don't even expect non-hostility. *You listen to them*. Including when they say they don't wanna get into it. Including when they show that they don't respect you, and when they lash out at you in ways you don't deserve. These things are signals, and their meaning is important, when you can't justify counter-bidding.

This is the weakest influence you can have, and at times it will also be the strongest you can justify. You might not want to engage, if this is all you can do. If you do, this is what you can always do. And often, this is enough.¹⁰

In short,

Theoretically clean:

Category	To bid hard	To request reconsideration of your bid	To request reconsideration of their bid	To accept	If you're wrong	If you're dishonest
Attention	"Am I ready to be wrong?"	"Am I missing something?"	"Are you sure you're not missing something? Are you ready to be wrong?"	Listen	Surprise	No surprise
Respect	"Am I ready to be humbled?"	"Am I taking your perspective seriously enough?"	"Are you sure you're taking my perspective seriously enough? Are you ready to be humbled?"	Show respect	Humbling	Humiliation
Security	"Am I ready to be vulnerable?"	"Am I being too threatening?"	"Are you being too threatening? Are you ready	Walk on eggshells	Sorrow	Excuses

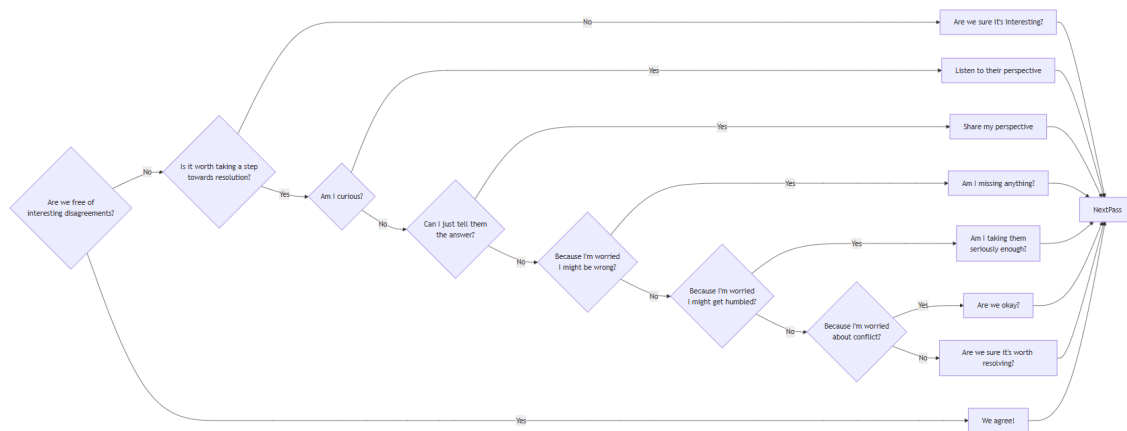
Category	To bid hard	To request reconsideration of your bid	To request reconsideration of their bid to be vulnerable?"	To accept	If you're wrong	If you're dishonest
----------	-------------	--	--	-----------	-----------------	---------------------

Or in more natural language:

Category	To bid hard	To request reconsideration of your bid	To request reconsideration of their bid	To accept	If you're wrong	If you're dishonest
Attention	"What if I'm wrong?"	"Am I missing something?"	"Are you sure you're not missing something? What if you're wrong?"	Listen	Surprise	No surprise
Respect	"What if they laugh at me?"	"Am I taking you seriously enough? Am I teasing too much?"	"Do you think it's possible that I deserve a bit more credit than that? Are you aware of how you're coming across?"	Show respect	Humbling	Humiliation
Security	"What if there's conflict? Am I bracing against anything? Is this <i>purely</i> fun for me?"	"Are you okay?"	"Do I deserve this? Will you forgive me?"	Walk on eggshells	Sorrow	Excuses

Navigating by these principles doesn't fit well into a simple flow chart. The actions we take evidence all three layers simultaneously, and therefore skillful navigation will take into account and act on all three layers simultaneously. We also split our attention, and go in multiple directions at once. This can't be easily represented as a set of discrete choices doing one thing at a time.

Nevertheless, when I notice a disagreement that I don't anticipate I will skillfully navigate without explicit thought¹¹, the mental moves I explicitly go through to help get myself on track roughly follow the following flow chart.¹²



In other words, this is my attempt to make explicit the steps required to orient to the things that are important for finding truth in a world where other people's behaviors and models are themselves evidence about what's true. It's an unpacking of the "Orientation" step of the "OODA loop". Each time through the loop our focus can shift to the new most important disagreement, which may or may not be the same disagreement.

For example, we might shift from "The place we should go for dinner is X" to "The most important consideration is whether they can accommodate Bob's peanut allergy" to "Does Bob actually *have* a peanut allergy!?" -- to "This can all wait; the building is on fire". This ability to shift object of attention each iteration is a feature, as it allows us to track the most currently relevant facet of divergence according to our entire ability to discern, and to respond appropriately.

In a different situation, the initial disagreement might be "This person thinks X, and I think that's wrong", but after orienting to this disagreement we might notice that we're flinching away from making a strong bid for attention because we're afraid of being wrong -- and we might think *this* is unnecessary and inappropriate relative to the strength of our evidence. If so, we can reorient to

this new internal disagreement and face the question "What if I'm wrong?" before returning to "This person thinks X" once we resolve that.

There's no requirement to say anything at each run through the loop, and when we can afford the time we can often get better results by anticipating results and iterating internally for a few passes before deciding on what to actually say. For example [when the drunk guy said he was going to drive home](#) I *thought* "That's definitely not right", but I didn't actually say anything until I had looped around enough to resolve the internal disagreements and notice that I was definitely curious and didn't understand his experience well enough to comment. In other cases though, sometimes I *will* just say things to probe and see what happens. I can make guesses about whether they'd get insecure, or laugh, or listen... or I can just say it and find out. Sometimes that's worth it too.

Aumann agreement actually works -- if you behave rationally

We normally think of [Aumann agreement](#) as not actually working in real life, which is because the prerequisites often don't exist in real life. We don't always have two people that are updating on the evidence of each other's beliefs. Therefore we also don't always have *common knowledge* that both people are updating rather than flinching into insecurity.

We *can* though. [And do](#). We can even practice Aumann agreement even with babies and toddlers.

The important part of pulling off Aumann agreement is that you stay honest and engaged with feedback. The moment I try to bullshit a bid, you can smell it. So you won't update, because it's not evidence of what I'm claiming. Neither will I anymore, because now your lack of updating doesn't mean anything either.

But what if we stay engaged?

When my baby girl cries, sometimes there's a real problem, but sometimes I get the sense that she's actually fine. The former bothers me, because I want her to be okay. The latter does not, because by definition there isn't a real problem in those cases. When I suspect the latter, I can be *both* empathetic and in a playful mood. So I'll smile, and I'll tickle her, and I'll expect her to smile back.

What happens next?

If I was right, she might soften, or break. It's really fun watching kids *try* to cry, and be unable to help themselves from smiling and calming down. That happens sometimes.

Other times, it won't help. At least not immediately. Sometimes she'll cry *harder*. What happens then?

Well, am I still empathetic and playful? Or did she successfully poop my party?

Depending on the situation, I might double down, she might double down, or both, or neither. Sometimes I think she's bluffing extra hard and I'll keep tickling, and she'll turn around. "I think not okay!" "I think okay :)" "I think not okay!" "I think okay :)" "Okay, I agree". Just another iteration of the Aumann agreement game.

In the cases where my initial suspicions turn out false, I start out "Aw, you're okay!", then soften to "Are you okay?" and turn around to "What's wrong, baby girl?". "I think not okay!" "I think okay :)" "I think not okay!" "Okay, I agree".

Aumann agreement works just fine. To pull it off, maintain the prerequisites. Keep close attachment to reality. At each moment, watch your confidences and beliefs and moods shift underneath you. Find out whose footing was more stable by whose stance changes in response to evidence of disagreement first.

Truth continues to exist, even when difficult to pin down

This sequence opened with an example of an insecure girl not wanting her makeup to wash off, and her friends disagreeing that this was important. It often feels like these things are about "differences in *values*, not disagreement about facts". But the fact is that no one would judge her, and once she recognized this to be true she was able to enjoy swimming and see first hand that her worry had simply been incorrect.

I then used the example of the guy in overwhelming pain due to nerve damage. This one also doesn't seem like "It's about disagreement over factual matters" but rather "It's about physical sensations!". Yet he ended up experiencing quite significant improvement quite quickly, without change in physical sensations.¹³ Because he had recognized different things to be true.

My friend with her fear of heights, same thing. Once again, what looked like "emotions, not facts" turned out to be a disagreement over facts made persistent by insecurity driven flinches from reality.



Chris Lakin ✓
@ChrisChipMonk



“why are emotions so illogical?”

emotions **are** logical you're just bad at logic

10:17 AM · Feb 16, 2025 · 137K Views



74



290



1.7K



310



These aren't cherry picked "coincidences" that happened to be about facts, only by chance looking like other situations that "really are" about things other than fact. The road marked by continued engagement with reality *invariably* leads to agreement over bets. The paths I took in resolving these things weren't predetermined because I did not know how they would respond to anything, but the navigational strategy of looking at new evidence as it arises and redirecting to what matters was predetermined. With my friend's fear of heights, I didn't know whether she'd leave saying "I guess that *is* dangerous" or "I guess I'm not scared no more", but I knew it'd be one or the other -- or, if not, that it would be because we failed to reach the end of that road.

This is because at any step, we're either in disagreement over what's true, or we're not in disagreement.

For example, you might say "But what if people *a little bit* judged the girl when her makeup washed off, and you just don't think it matters but she did!?". Well, then in that case, the disagreement would be over whether that little bit of judgement matters. I might hypothetically believe that it just makes the judger an idiot, and she might think it's the end of the world. These are factual things though. If we can summon the security to face reality, we can *find out* whether the world ends. And whether the judger is an idiot.

If the world ends we can't really continue, so let's assume it turns out that the judger is actually an idiot. Or, more precisely, someone weaponizing and externalizing their own insecurity to protect themselves from uncomfortable truths, and in the process pushing ideas that are obviously wrong *even to them*. Even once we realize that this is going on, she might not want to go swimming! She might object "I don't like seeing people as idiots!", which sounds like a "value" thing again.

But why would she not like it, if not that it's "bad" in some way? Maybe she's right! If I still think the right play would be to write them off as "idiots", then we still have a disagreement to face: when people do things that are predictably-to-them stupid in a way that attempts to force

other people to pay the consequences for their own insecurities, is it better to dismiss them as 'idiots', or to not? These are all questions that could conceivably be answered, if it were to get that far.

Things don't bottom out at "Values difference!", because even if it's hard to see how "values" cache out in "truths", we still have to decide which values to prioritize, and there are better and worse ways to prioritize things.

If we run into a "Values difference!" of "Truth vs compassion" for example, then there's still a "Way that things will play out if you prioritize truth" and a different "Way that things will play out if you prioritize compassion". Sure, in the short term, watching someone's feelings get hurt by a tactless pushing of 'truth' won't change any minds, but that's just because the disagreement over implicit facts lies further in the future. If the community decays because of insufficient contact with truth, and people lose everyone they love, even the "compassion" favoring people will have something they can't ignore. If the community becomes nothing but fighting because shutting out truths relating to the value of compassion made things *immediately* fall apart, to never improve, then the "truth" favoring people won't be able to hold onto their perspective without shutting out truth themselves.

Whenever we feel like "We're both looking at the same reality, and disagreeing over values", what that shows us isn't that truth "doesn't exist", or "is relative", but that we aren't yet aware of what implicit differences we're in disagreement over. One side might see white lies as deeply *wrong*, and not know why. The other side might see using 'truth' as an excuse to be mean as *obviously* wrong, a priori, because compassion is just what matters. But we can always dig one step deeper, and ask: *Why* is compassion what matters? *Why* is it wrong to white lie? What's the harm, in either case?

And when we look, we can start to notice. The harm in white lies is that it breaks contact with reality, increases the risk of crashing into rocks, and of not even noticing that you're sinking and that this could be avoided. The harm in pushing truth without compassion is that the subtle insecurities that lead you to do it without compassion also give hints that your implicit attitudes aren't knowably true either -- and [those will sink you in very specific ways too](#). Careful allegiance to truth gets you compassion as well, when it's true that compassion is good. Like when someone is in such a rough place that they judge risking a car wreck as better than staying at a party, and the implicit worldviews motivating "That's dumb!" turn out to be false. Or when a girl at a music festival is being a turd because of whatever reason, but whatever reason that actually exists and is sufficient to make responding to her with anger or contempt *wrong*.¹⁴

The road is always there. The only question is whether you make the decision to stop engaging before you get there.

The Default Failure is Arrogance

If we meta-disagree in the direction of thinking each other's object level perspective is worth more than our own, then things resolve themselves. We're both willing to listen, including to the part where the other side thinks we have something of value. Therefore, persistent disagreements are always errors on the other side; at least one party is arrogantly shutting out evidence contradicting their perspectives.

We then get to thinking "It's definitely me that's right. But they're so wrong and irrational that I can't just say that, so I have to do things to fix that" and end up puzzled how to solve these problems with the *other* person's thinking.

When I notice myself a bit confused, thinking that I know what the right answer [probably] is yet not immediately knowing what to say/do to resolve the disagreement, the place I like to start is to notice: if I'm definitely so right and justified and rational, *I can just say what's true and that is enough*. When I know I'm right, my confidence will shine through. I won't shy away from bidding from attention, even if people might laugh. I won't have my feelings hurt when they laugh, nor will I fear conflict. After all, what I'm saying is true.

When I notice that I *can't* stand so strong without doubts, and cannot actually anticipate that people will come to agree, that means I'm not as justified as I've been telling myself *and I know it*. So might as well start acting like it.¹⁵

This is why there's no "debugging section" to this sequence. The *entire sequence* is the "debugging section" to the idea that if we're so rational and justified and right, we can just say what's true and expect people to believe us, because they can see that we couldn't face reality like this and maintain our belief if it weren't likely to be true. We can do Aumann agreement *with a baby* if we maintain contact with reality. Simply pointing at reality unflinchingly *is enough* -- when we can pull it off.¹⁶

That's why I can basically say "Turns out you can just decide not to swell injuries". Or simply "Let me know when you're ready. Okay, now go into hypnosis" to a person who starts off thinking they "can't be hypnotized" because someone actually tried and failed to hypnotize them. That's not always the quickest path to understanding, which is why I broke it down into pieces the first time I wanted to help my wife change the amount of blood flow here hands were getting -- and didn't just say "[If your hands are cold but your core is warm,] Then send more blood to your hands, right?" until after she had that experience.

The fact that we don't *have to* give that more detailed explanation, and can simply hold the line, is usually not important as "the right play" but it can be very useful as a starting place to debug what we'd like to happen.

From there, we can notice *if* that would work, and if it wouldn't *where* it would go wrong. What will their response be? Where do we feel pressure to fold first? What if we fold preemptively *because we notice we would be overextending not to*? Humbling ourselves before getting humiliated. Chilling on the provocation before getting attacked. Etc.

It gives us a concrete way to 1) get maximally simple and quick results, or else 2) find out *exactly* where we don't yet deserve them.

Do I notice that I'm afraid they might actually believe me?

Do I notice that they might not take me seriously, and I'm afraid they might turn out to be *right* not to?

Do I notice that I anticipate conflict, and that I'm not comfortable enough to stay open and vulnerable, knowing that my position is safe and solid?

The pool intervention, explained

Let us now return to [the girl not wanting her makeup to wash off in the pool](#). How did I generate those moves, and know that they would lead to a good outcome?

The perspective she was expressing seemed kinda... dumb. Like, no, that's not real. I don't think, at least. I have a hard time taking it seriously. That's what seemed real to me, at the start.

Did I know *for sure*? If I were to have a magic programming interface to put new beliefs in her brain unilaterally, would I do it? No. I wasn't *that* sure. There are lots of ways that even explicit delusions can be load bearing¹⁷. There's way too many ways for that to be a bad bet that I prefer to be [guided by the beauty of my weapons](#); I want her to *agree* with me that my perspective is the right one, and for good solid reflectively stable reason, if she's going to take it on as her own.

Would she do that, if I simply asserted it? Clearly no. Everyone can tell that this is a "No". Her friends had tried, even. Poorly, but they tried. I don't anticipate her simply accepting it if I asserted with an expectation of her accepting my words as truth. I anticipate her pulling away out of insecurity. Which *seems fair* to me, actually. Because "Who the heck is this guy, and why does he think I should believe what he wants me to believe?". I agree.

I don't want to talk if that's the effect it will have. If I want to engage, and I don't want to talk, my option is to listen. I can attend to what *she* is pointing our attention at, which is the perception that if her makeup washes off it would be bad. Presumably because the makeup would no longer be there to do makeup things. So "You really don't want people to see you without your makeup on, huh?".

That's pretty validating though. Do I actually see those feelings as that valid? Am I really curious to know more about what her feelings are on the matter, or do I already know? I kinda feel like I already know where this is going. I'm not honestly that curious, and I don't really take her protestations very seriously. So what if I signal that? What if I reflect back her statement -- *granting* her attention -- but signal that I don't exactly take it completely seriously? How can I do that?

Well, I can say what I have to say in order to keep rapport, and let the humor I find in it shine through. Is it *genuinely* funny to me? Yep. When I check my anticipation of how this bid for respect goes, do I anticipate her holding her ground and saying "Thank you! Some people don't recognize how important this is!"? I don't. That would genuinely surprise me. It *could happen*, and if it did I would have humble pie that I would need to eat if I were going to stay engaged with truth. Was I ready to eat it if it came to that? Was I willing to take that risk? Not if I was outright *laughing* at her while saying it, no. But if I'm merely exaggerating the parts I don't expect her to agree with, and leaving subtle hints that I don't really buy it? Yeah, I can eat that much humble pie. I'll take that bet.

But might it upset her? Might bidding that boldly push her into considering ideas that felt more threatening than she was confident she could process into *better* results? It might. Am I willing to risk that? Well, again, depends on how I go about it. I didn't want to *leave* her upset. I didn't want to come off like a jerk to people whose perspectives I respect (e.g. mine, my wife's, the girl's, in the likely case that she were to turn out to be reasonable). But is that really likely? Nah, I'm actually not judging her. So long as I'm careful to only reflect her own judgments back at her I can easily side step any offense; I know how to be that careful. She just can't stay upset because I can say with an arbitrarily straight face that I'm not calling her ugly, just giving her the respect she deserves (again, because it's true, and I *mean it*). She might get upset, *and that's fine* -- because her upset would be wrong and I can prove it to her own satisfaction. Her upset would only be temporary, and lead to a better place. Could I get *this* wrong? Absolutely. But if I did, it would genuinely surprise me, and I would be genuinely sorry. So I could actually *mean it* if I were to owe her an apology, and I don't anticipate any reasonable chance of her not accepting the apology as sincere and sufficient.

So I knew my bets.

I'm not gonna tell her she's wrong. I'm going to give her the attention respect and security she asks for. At the same time, I'm going to play my bids that I don't *entirely* think this is the right answer, or justified, and that I *do* expect she'll find that I'm on her side. Because I didn't think her answers were right, or very well justified, and I did think she'd realize that I'm on her side. Because I was, and wasn't going to flinch from engagement.

So I played those bets, and was proven correct. Which earned me the security and respect needed for her to take my sincere perspective as meaningful enough to both ask for it and contemplate it.

Once she asked, it was clear that she actually wanted to consider my stance, so it felt worth giving. So I gave it. And she considered it. And all things considered, it seemed like I was probably right -- so she knew her new bet too, and going swimming seems to have paid off.

I opened this sequence by posing this challenge and stating that it could be solved by simply showing the truth and justifying it.

Now that you've seen the framework, and now that you've seen the exact process of checking which bets were justified, do you see it?

Do you see how those lines are specified *almost uniquely* given my epistemic state at the time?¹⁸

Building security on purpose/Justifying security

A lot of options open up when you're secure. You can make fun of strangers in ways that they'll enjoy and feel safe with, only for them to ask for your input on things that you would never have thought they'd listen to.

Problems tend to dissolve when there's no barrier to looking at reality. But security is hard earned.

In the "Clarity" post, I implicitly conflated "security" with "safety" a little bit. Yes, it's easier to act with security when a toddler is throwing a tantrum than it is when a man with a knife is throwing a tantrum, and that's because it's objectively safer. At the same time, [security, on a fundamental level, is an ability to compute acceptable responses to the new information we're presented with](#). If you anticipate that facing reality will drive you to harmful conclusions, then you will not feel secure. If you anticipate that facing reality is the best you can do, you'll face the damn reality. Even if it remains objectively unsafe.

If that violent man with a knife is *definitely* going to find you, and he will *definitely* try to kill you, and you *definitely* can't run... then when this is crystal clear to you, it is no longer tempting to cower or squeal in fear. You stand up and fight with what you can, even though it's unsafe, because it's the clearly the *least* bad thing you can do. A good *definition* of courage, is the embodied recognition that the best course of action remains scary as fuck.

So how we build security on purpose? Well, first by staying *safe*, if we can. If we can't be threatened, then it's obviously going to be okay to face reality.

But when we can't ensure safety and good outcomes? We can think through what to do with the information we might be confronted with, which we're choosing to avoid. "What if I get hurt!?!?". What if they think I'm dumb?

What if?

Like, non-rhetorically. What if?

Can we avoid it?

If not, what *would* we want to do about it, should that turn out to be the case?

Well, maybe we want to fight back with the best weapon we have, even though it's a chair. Maybe they'll think we're dumb and that's unfortunate, but at least we know they're wrong. Or maybe they *wouldn't* be wrong, so, uh... I guess figure out what to do about the possibility that we're dumb?

Once we have answers there, it's not "not-scary", exactly, because you could still die. Or get made fun of, which might feel worse. But it won't be paralyzing, because the right answer will be clear.

Security is built by justifying security. By attending to the question "What *would* be the appropriate way to respond to the threatening information?", and settling on an acceptable answer. In other words, "[Leaving lines of retreat](#)".

Oh no, people might not laugh if I make that joke. People *might* laugh, if I say that for serious. What if they do? Does it mean I'm wrong? Do I want to humble myself or play 'double or nothing'? Not always, but often these can be answerable in the moment, with just a flicker of awareness that there's a question to be answered. In the jacuzzi most of those questions had intuitively obvious answers, but I had enough insecurity that I would have ordinarily kept my mouth shut. What if she gets offended!? What if I misjudged, and this is a more serious issue than I've assumed?

I'm not judging her, *obviously*. That's why I'm not accusing her of silliness, and instead gently hinting. I can eat this much humble pie, and I'm prepared to give an apology sincere enough that she will accept it. This much was obvious in a flicker, and I did run these checks explicitly before opening my mouth to tease her. This sequence is both a descriptive explanation of what you already do, and a set of scaffolding that you can use to make sure the things you find most compelling in the first person cohere with the things that you think *should* compel you -- from the third person perspective. When you find yourself displeased with how things are going, you can think back to the key questions, and notice any discrepancies. If you think you're right and worth listening to, but you're *not* bidding for attention and respect because it feels bad, then is it

because you're bullshitting yourself or because you're insecure unnecessarily? When you face and answer the "What if!?" questions, what happens to your willingness to stick your neck out and bet? What happens to your third person stance on how right and justified you are?

It can also act as scaffolding to make sure other people agree in the first person as well. Often, the biggest obstacle isn't knowing that we're right, it's the other person knowing how to recognize "losing" the argument to be winning in life. Stepping into their shoes and retaining security does serve to bid for their security in a content-less Aumann Agreement sort of way, but retaining security requires *you* to know how they come out ahead by listening.

If the irrational butt head doesn't want to listen even though you're definitely right because it threatens his ego, then how *is* taking that blow a winning move, for him? How would *you* deal with that information, and know you're coming out ahead for it? Because if listening to you will cause him to lose, he's making the *right* decision by blowing you off! And if you know the answer, you can also articulate it. Or highlight it. You're allowed to embody that knowledge and let it flow downhill in all ways, not *just* through body language.

Think things through enough that you know what you'd do with that information -- or just wing it, and have faith in your ability to figure *something* out -- and you'll have security. And you'll be able to invite them to join you it by bidding credibly.

When you're not happy with the results you're getting because you're flinching left and right, this is your homework to do. What *would* happen, if you were to not flinch? And what might you want to do about it, if the thing you're flinching from were to become real?

Rationality is simple; preserving it is not

When learning to shoot, the way to hit the target -- rather than flinching -- is remarkably simple: you try to hit the target. That's literally all it is. My "novel" solution which fixes flinches orders of magnitude faster than regularly practiced methods is simply "Make sure to not do the wrong thing, try to do the right thing".

The fix itself takes a few seconds. The explanation of how to fix it takes a few minutes.

Why does it take minutes and hundreds of words to convey "Make sure to not do the wrong thing, try to do the right thing"?

Why is that a thing that even needs to be said at all?

Let's look closely, starting at the base layer:

To hit the target you have to aim at the target

To aim at the target you have to decide to aim at the target

To decide to aim at the target, you have to not expect that you can't decide these things.

This last one highlights the difficulty.

Once you've formed the belief that you can't decide these things, that belief gets in the way of doing the things you'd be doing "automatically". It is this belief that breaks your contact with reality, and prevents you regulating towards what you want in optimal fashion.

Where do these beliefs come from?

Well, look at what they *do*. These beliefs serve to stop you from *trying*. You can't decide these things, so *don't try*. If you try you will fail. You will experience prediction error! Suffering! Dukkha! We don't want that!

These beliefs serve to protect you from facing the pain of reality invalidating your fantasies. But don't we want that? Isn't that how we summon the motivation to *act*? Isn't that how we know to push against whichever are the weakest obstacles in our way so that we can reshape the world to *get what we want*? Haven't I been hammering on the fact that our feelings say things about reality, and we want actual reality to be a certain way? That we don't give a shit about "feelings" for the sake of feelings?

Well... yes.

So again, what's going on?

Again, pain illuminates. Bump your toe, and it's easy; it hurts, you realize "Oops, there's a rock there, I'll be more careful". Then you move on with your life. Have your nerves fried in a medical situation gone wrong, and it's not so easy. Instead of "Oops, here's the obvious next move", it's "HOLY HELL WHAT THE HECK IS GOING ON!?!?". *Everything* feels urgent. Which things are? We don't have space to figure out. We cannot contain the entirety of the situation at once. We can only productively give so many fucks at a time, and getting blasted in the face with a firehose of information does not lead to quicker faster decisions.

So we look to back away. Rationally, but often without self awareness.

In doing so, we enter an unstable situation.

The concept of "I can't force myself to do what I think would be the right thing to do" lives very near to "I can't decide". They are distinct, and only one of them is true, but unless you have crystal clear vision it is easy to genuinely mistake one for the other. Which one is going to be the

more attractive answer -- which one is going to feel like it gives you more of what you want, in the moment -- if there's ever any little bit of pressure to face reality when you feel like maybe you can't handle it? What happens next, when the response to "You have to face reality" is met with "I can't. And I am, already"? It's not a *decrease* in the pressure to confabulate reasons for failure.

Once you have motive to say "I can't. It's not so simple", you're at very high risk of believing it.

And once you believe it, how can you get out of the trap?

Well, you can be taken by the hand, and *led* out, one issue at a time. When you're hurt, and try to fight off tears not knowing what to do with the information you're blasted with, sometimes all it takes is for someone to squat down to your level, join you in your suffering, and show by example that it's possible to face it. This can take very few words, or no words. Experience this enough, and you might expect to handle more things by default. And start sitting with others in return. Or, at least, contributing less to mutually prickly interactions and averting some melt downs. Which also contributes to those around you learning that it's safe to lower their spears and shields a bit, on the margin.

At least, until you're faced with challenging enough situations that the neutrons begin multiplying again. Without an awareness of these dynamics, and a recognition of the way out, group dynamics will blow with the wind, with no one at the wheel with the recognition of how to stop the descent when things go poorly.

The point of the words is to paint a picture of the problem. See "I can't" for what it is, and it's no longer so compelling. Because you see the gaps, and through them, the way out -- back to the difficult question.

For a localized flinch such as when shooting a handgun, this may only take a couple hundred words.

For chronic pain, it may take a few thousand.

For the fully general case, needed to recognize and navigate every facet that "Pressure to face more reality than known-productive generates flinches which get confused for beliefs and reality" expands into... such that swelling injuries feels like a choice... such that you can see the flinches in chronic pain and help lead people out of their suffering... such that when your community flinches from reality in ways that it is both genuinely and motivatedly blind to, you can see the structure that reinforces it and how to nudge things in a productive direction...

I don't think 70k is enough.

Even if you've read every word of this sequence, I doubt that swelling injuries *feels* like a choice. Or that pain feels like "just information we care about". Or that you'll see willful arrogance every time it's there, especially in yourself, when "It's an honest disagreement and/or misunderstanding" is still a socially acceptable line. Experiences need time to add up, and alternate frameworks filled in. Reality is still a fire hose. We all have to pull away, at times, and regulate out exposure. And once it comes time to reengage, on the margins, it's not always easy to find where to start.

But maybe now the cracks are visible. Or, the possibility that *there exist* cracks, is visible.

If I have been successful, then you will start to notice. The insecurities you harbor will feel more obvious, more understandable and *unavoidable* rather than simple moral failures -- and simultaneously more *optional* -- perhaps wise, perhaps not, and up to you to tell.

And so when it comes time to reengage, on the margin, you'll know where to look, and how to tug at the seams.

When you find yourself flinching, you'll know you can look at the relevant uncertainties, and aim to resolve them.

Like you already do, at your best. Like we *all* already do, at our best.

Just, with a better map for diagnosing and noticing when we aren't at our best, what the ceiling is, and how to get back to where we can see reality for what it is and enjoy playing with what's real.

"Irrationality" is fundamentally a process of insecurity -- of rationally but perhaps not reflectively-consistently deciding that we can't handle the truth. Rationality is finding our way back to the default state. It's opening our eyes and looking, literally and metaphorically.¹⁹

Notes

1. And more. There are two more posts coming on limitations and failure. ↩

2. It happens. Complete honesty is actually a really high bar to manage, and it requires self monitoring on these things quite closely.

The take away, if you catch yourself bidding dishonestly, isn't "Oh no, I'm a no good liar", it's "Well *that* was a mistake. Oops". ↩

3. In the [chronic pain example](#) I didn't ask explicitly, but I knew of a bunch of things that I would appear to be missing if I did not address them, so I addressed this question preemptively by showing all the things that I'm *not* missing, which would ordinarily turn people away from the stance I was taking. ↩

4. In the "irrational fear of heights" example, when I kept redirecting her to "How do you actually know you're safe up there? Or *do you?*", that is *exactly* this. ↵
5. An example of this is where I was able to help my friend reexperience hypnosis despite her laughter and claims that she couldn't. ↵
6. In the jacuzzi situation, the sincerity behind the question "I just think if you're worried, it's probably for a good reason... right?" was actively keeping tabs on "Am I taking your perspective seriously enough?", even as I kinda made fun of it. ↵
7. Similarly, when teasing my cousins girlfriend about being a diva and helping her get back on the horse dirt bike, visibly attending to "Are you [still] okay?" is a big part of what made that work -- though that attention didn't have to be verbalized. ↵
8. The 'one word pain cure' of "Hurts?" isn't quite asking *if* he's okay, but it is asking what *justifies* his belief of not being okay. It still came from him looking less okay than he seemed to me, and it's still asking "Okay, maybe you're not okay. Is that true?". ↵
9. Note that this turns out to be *exactly* what I said to the woman at the music festival: "Are you okay? Will you forgive me?". At the time of doing that and even writing about it, I had no idea it'd distill down this cleanly, but it's quite neat that it did. ↵
10. The situation with the drunk guy at the party is an example of where I had no better option, despite someone with a more privileged epistemic state coming in and demonstrating a more effective response. Still, it was enough, and when I saw him again a year later he made a point to express appreciation. ↵
11. E.g. someone tells me of their chronic pain, and it's not obvious what to do about it. ↵
12. And again, this is what we *all* do in one way or another when we're navigating skillfully enough that we can just expect things to work. It only needs to become explicit when we're struggling to hold that expectation. ↵
13. He reached out to me again recently, almost a year later, and his pain killer use for this issue has essentially dropped to zero. He said he "basically only takes pain killers for migraines". ↵
14. Truth and compassion aren't actually at odds, at a fundamental level. It's just harder to figure out how to satisfy both simultaneously, at times. Any high effort resolution on this one will do better by finding ways of having both. ↵
15. There's an important difference between statements like " $2+2=4$ " and statements like "You will agree and acknowledge that $2+2=4$ ". If you can justify the latter, then you will be able to anticipate they will agree. However, if you can only justify the former, that's not necessarily the case.

If there's pressure on people to falsify their beliefs, and it's genuinely not in their best interest to believe or be seen believing " $2+2=4$ ", being justifiably correct about " $2+2=4$ " will enable you to hold your ground as you bid for attention and respect, but security will be hard. And as a result, you will be able to anticipate that you can hold your ground and bid hard without feeling pressure to fold, but that *they* will feel pressure to fold, and will eventually be forced to retreat into insecurity -- or else acknowledge the truth.

In these cases, where you're focusing on " $2+2=4$ ", and not " $2+2=4$, and it's okay to say that because it's true", you're *not trying* to resolve disagreements. Because you're not trying to justify the idea that they should agree. In these cases, you're staking a claim and enforcing a filter: "Engage with truth, or your cognitive dissonance will be exposed to you and to any audience that might see".

Whether or not this is a good thing in any given situation is a separate question. ↵

16. And when the interaction doesn't get interrupted before it can run its course, I guess. ↵

17. As talked about in [expectation=intention=setpoint](#). ↩

18. What I hope to convey is more than just "I see how those lines can come from what you can justify". I hope to make it clear that nothing else I could have done could have been as good, in expectation, given what I knew.

Pure empathy without the playful challenge might have gotten her in the pool. I don't mean to say that it couldn't have worked. Same with a more overt and direct approach. I potentially could have said something like "Hey, you. Yeah you. Listen. I have something to tell you that you know to be true in your heart of hearts, yet which you have your reasons for not wanting to listen to. I promise you that if you listen to your heart, and these words, you will be safe, okay, and happy you did. Do you understand? Ready? Okay, great. *It doesn't matter what people here think of your looks. If anyone were to get judgy about it that'd make them a jerk, and that would be their problem. You're fine. Enjoy your day ;)*". That *might* have gotten her in the pool too.

I'm saying that I wasn't confident enough to justify the latter, and I was more than confident enough to justify the former.

If I had said the latter, I might have ended up with my expectations falsified when she didn't believe in her heart of hearts what I said she would -- and then I'd be left a fish out of water. She might have cut me off -- for whatever reason -- and I wouldn't have a next move because I would have been caught overextending. It's hard to expect people to take you seriously as someone whose words mean anything after you've just demonstrated that you're out of touch with reality.

If I had said the former, I would be leaving influence on the table. I would be lying by omission, dishonestly conveying that her situation was too scary to even think about. She might have still got in the pool, but she might not have. She'd have noticed that my inability to laugh is evidence that her insecurity might be justified, and as a result might have held onto it longer than she needed to.

Both might have worked, but neither are optimal relative to the confidences I actually could justify in that moment. ↩

19. If Valentine's post on [Kensho](#) seemed confused or confusing, you might want to try rereading it now. ↩

Part: Limits and Failure

Finally, we close with an exploration of the limits and failure.

On physiological limits of sense making

The last post, "[Putting it all together](#)" served to close the loop and complete the framework. From here, we just have to tie up some loose ends and demarcate the limitations of what can be done both physiologically and practically. This post discusses the former.

When the "disagreement" frame falls apart

Common perspective is that unreasonable people aren't updating on evidence when we know they should. Irrational fear is about our own brains somehow not doing what we know they should. Pain is about nociceptors firing. Swelling is just a thing that happens. Depression and bipolar disorder are about chemical imbalances. Alzheimer's is about physical brain degradation. Phineas Gage's change in temperament was because he got a railroad stake through the brain, and gunshots to the head kill by physical damage -- not by convincing people it's no longer worth beating their hearts, breathing, and living life.

By exploring the framing of "disagreements" and the fundamental requirements for navigating them successfully assuming the underlying neurology *is* updating properly on new information, we've noticed that common perspective overstates the physical limitations and understates the degree to which we just don't know what's right. Yet clearly by the time we get to the end of that list, "physical malfunction" becomes a significant part of the story.

Now that we've seen how to navigate disagreements when the underlying hardware works well enough to be abstracted away, we're equipped to notice when it *doesn't*, and explore how that changes the analysis.

When a system can't learn

Let's return to my temperature controller, to start with.

Back in "[Expectation = intention = setpoint](#)" I quipped that even temperature controllers can have insecurity issues. My heat treat oven temperature controller self censors if it ever tries to expect to put out more than about 40% of maximum power. It embodies the expectation that the outcome won't be better if it engages with the full extent of prediction error, and limits its ability to address that prediction error as a result.

What if we wanted to teach it that it's safe to be more secure? What if I changed the elements out for elements that can handle the full output? I can and do change the code to update such expectations, but there's nothing you can do from the outside to fix this. If you bypass the

controller and force full power, it won't notice. Even if it had a current sensor and was closing the loop, there's no code there detecting whether it's still working, let alone code reasoning from here to what kinds of limits are likely safe. It's just hard coded, and if you provide it valid evidence that what it believes to be dangerous is in fact safe, *it won't respond*.¹

A person might give you the cold shoulder, or put on a poker face. Or they might let their eyes glaze over when your ramblings have no relevance to them. The temperature controller won't "ignore" you, *it won't sense you*. If you bring out a sledgehammer and threaten to destroy it, the temperature controller won't so much as blink. It will keep doing *exactly* as it was doing before.

Do you ever see such clear failure to respond to *obviously* relevant stimuli in people?

Ever see a *person* fail to respond to a sledge hammer raised over their head? That's one way to get a response out of someone who is trying to ignore you.

Or fail to respond *in any way* to getting blatantly laughed at? Not to stand their ground, or get defensive, or to try to ignore you, but to be in a position to *actually care*, and somehow *not notice*? I have. But the case that comes to mind involves talking with someone who was quite drunk... and on LSD... and, it turns out, asleep. So I think it's fair to hypothesize that his hardware wasn't very functional at processing information at that time. And, now that I think of it, he probably didn't care too much either.

When someone's brain is doing something dumb, it's often not easy to get them to agree. Or to hear you out. And people will often miss, or misinterpret, or choose to ignore signals that they're wrong, at first. But usually, if you have some credible reason to think their brain is doing the wrong thing, you can get them to *notice*. Their response might be a defensive flinch, or a conspicuous lack of active response where there otherwise would be, but even that is responding to the stimulus with a change in behavior. It's not often that you can escalate things all the way and have people act *exactly the same* as if they don't see the evidence, the way my temperature controller acts like it doesn't see the evidence.

Usually, what happens when you tell people "You're definitely wrong" is that they *respond* to that information in some way -- and either they agree or you have a disagreement to sort out. Generally people are functionally intelligent enough to notice the disagreements, but that doesn't necessarily mean their hardware is functional enough to allow *resolution* of the disagreement.

What would it have looked like if *my friend* was *right* that her fear was irrational? If her fear was *actually* immune to evidence due to a brain tumor or something, how would we notice? What are the signs that a disagreement is actually unresolvable due to hardware limitations, rather than simply an inability to provide the appropriate evidence?

Take a moment to think about it for yourself.

And when you're ready, we can explore some actual examples.

Amnesia by force

Like the time my friend in high school came in to class and started telling me about how his head hurts, and his arm hurts, because he had been jumping on some cardboard boxes outside and fell off hitting his head and his arm. A minute later he started telling me again about how his head hurts, and his arm hurts, but this time he didn't know why. I told him he fell when jumping on some boxes, and he wouldn't believe me. He thought I was messing with him. That's a disagreement.

If his amnesia were purely the result of software, like "hypnotic amnesia" is, then in order to maintain the amnesia *he would have to find reasons to not look*. That's not what happened though. He didn't protest that "I *can't* remember! I have amnesia, remember!? You can't expect me to remember!". He didn't insist that he's "trying", while directing his attention towards me, or his backpack, or any other place that's unlikely to contain the memory. He didn't look like "This makes sense, I don't *expect* to remember what I did 5 minutes ago".

When I told him that I was serious, he looked *confused*. Like, "If that's true -- and I believe he's probably right -- then I would expect to remember that. *And that expectation is not being realized*". I could see him... well, I couldn't *quite* see him looking for the memory, but I could see him *not* focusing on why he didn't have to look, and *not* focusing on anything else. He sure looked like he was actually looking, and drawing a blank. It's like when you're fumbling for a word that's on the tip of your tongue, saying "Can I borrow your... ugh... what's that word. Hm..... OH! 'Calculator!'". He was visibly looking, only the memory never came. In other words, his behavior was better explained by a genuine hardware malfunction.

This is unsurprising given that he had just bonked his head on the concrete, but knowledge of this blow to his head isn't necessary in order to notice that *this doesn't look like software*. He wasn't just *saying* that he should be able to remember, he was actually aiming his attention in the directions needed to retrieve the memory *and it wasn't working*.

We were able to agree that he had amnesia due to a blow to the head, and that I'd walk him to the nurses office, but when it came to the memory retrieval his neurology just wasn't doing its job. There was a legitimate physical malfunction, *and* a disagreement resulting from that. Resolving the disagreement was necessary in order to do the appropriate thing for the context, but that doesn't automatically fix the physical damage or bring back the memory.

Fried nerves

The chronic pain example also has misfiring neurology. His nerves weren't *actually* sensing "everything", but that's how it felt because the nerves were physically damaged and misfiring constantly. The conversation I had with him didn't fix his nerves overnight so that instead of "feels like everything" it felt only like what was there. It just helped him contextualize *what these nerve signals meant*. His nerves weren't accurately mapping reality, but his brain was able to map *that*. Instead of expecting his nerves to map reality, and expecting this expectation to be realized (thereby driving him crazy when nothing was going as expected) he could notice that this expectation was going to fail -- at least for a time -- and make sense of what to do now that these nerves which were formerly part of his roughly coherent system were now more of an "external noise". The overnight benefit of that conversation was not in fixing the reality of his damaged neurology, but in helping him understand how best to navigate the new reality that he found himself in -- partially malfunctioning neurology and all.

Truly irrational fears of height, and transient PTSD

Returning now to my friend and her fear of heights, can you see what it would look like if her problem were indeed hardware?

If you want to take another moment to test your understanding, do so now.

...

...

Got your answer?

Having a fear that doesn't respond to evidence is like having nerves that don't calm down once you realize that you're not actually sensing "everything". If the nerves don't respond to reality, you stop responding to them. You stop acting like the fear means anything, once you realize that it doesn't. Okay, I'm afraid -- *so what. It's not dangerous*. It becomes *funny* that you're afraid, and the fear stops being bothersome. You just route around it, and depending on exactly where the disconnect is, it might even stop feeling like "fear" and start feeling more like "inappropriate excitement".

This is also what it feels like with functional neurology, by the way. It's just that with functional neurology it's transient, as the object level beliefs update to match the evidence. For a month or so after a particularly bad experience I would have bouts of seemingly content-less anxiety come up for seemingly no reason. But it didn't bother me because I knew the bad was over, so it was

just "Ooh, neat. Anyway..". It was just a thing my body was doing, and it didn't *mean* anything. Uncoincidentally, it stopped happening.

Notice that whenever you have some "irrational" behavior that *isn't* going away, *it's not funny*? Notice how if you try to laugh, it just brings up other feelings which conflict with that laughter? Feelings about how you *shouldn't* be feeling this way... as if you could feel differently... if only you were to do something about it that you're not doing"? It's theoretically possible to laugh at your own responses and route around it while they stick around -- if your hardware isn't working -- but I've never seen it happen.

"Chemical imbalances"

Okay, moving on to bigger scarier problems. What about depression and chemical imbalances?

"Chemical imbalance" alone doesn't say much. We're built on neurochemistry, so it's no surprise there are neurochemical correlates; it'd be a surprise if there *weren't*. The question isn't whether there are neurochemical correlates, the question is whether those neurochemical markers are tracking the evidence or not -- and that's a harder question to answer.

Sometimes people are more depressed seasonally, and bright lights help. Sometimes fixing nutrition elevates mood. That seems to suggest that the chemistry is tracking things other than "rational evidence of how good life is". I don't think that's the full story.

Say you feel bad all the time due to poor nutrition. You're unhappy, and find it hard to get motivated to do anything. Well, what do you expect? If you're in poor shape and don't have energy to do much, *of course* you can't find motivation to do much -- you don't have *resources* to do much, so that what tracking reality looks like. If you're unhappy, is it just because you *want* to have more energy to run around and do stuff? That sure sounds reasonable, and "let go, and be content to sit around with no energy" doesn't sound like a great solution, when compared to "get your nutritional problems figured out". There's nothing *blank* about "Why not be happy?", the way my friend's mind was blank when trying to remember how he hurt his head. It's just *this actually sucks*. The problem is *real*, and the beliefs are doing their job by pointing at it.

Continuing the same pattern, this "depression" is a disagreement coming from a real limitation. That's important to notice, because if you know that your depression stems from malnutrition and a real lack of physiological resources to do things with, then you know to see a nutritionist rather than a psychiatrist. And clear thinking keeps us from spiraling into unsupported catastrophizing, so things aren't worse than they have to be given our real limitations.

While that's clearly a part of it, and I've seen it, I don't get the sense that most depression is even this physiological. Certainly my own experience with it wasn't. Everyone likes to say "You can't just decide to not be depressed!", but it's factually untrue.

When I was in middle school, my mom noticed that I had seemed depressed, and asked if I was. Like my cousin with the fire poker, I noticed that my immediate answer wasn't what I wanted to stick with, and by the time I opened my mouth to say "Nope", it had become true. In that moment, I *decided* to not be depressed, and it worked.²

I don't talk about it much because normally it'd just come off like "I could do it so you should too!" which isn't actually correct, and it would take many words to convey the correct meaning. And because "Good for you. What do you want, a cookie?" -- and I don't like cookies. But after spending tens of thousands of words explaining exactly why "just deciding" is so hard yet actually possible, it's worth pointing out that yes, this applies to depression too. There's just a selection bias in that the cases that are that easy to solve don't tend to gather attention in the first place, and people aren't exactly incentivized to share their stories.³⁴

Similarly, "bipolar" looks a lot like what you'd expect when someone oscillates from flinching from the good to flinching from the bad, changing strategies for how to avoid prediction error out of insecurity as each direction becomes unsustainable. Most people would think to lift people up when they're depressed and "bring people down to earth" when manic, but if you know anything about control theory you'll know that this doesn't provide damping. If anything, you might be a bit slow to notice phase shifts, and end up contributing to the instability. Instead, if you look for the flinches from reality and orient to truth, you'll end up *leading* the mood, which provides damping. This means tearing people down even at their lowest, as they start to ignore the bad instead of the good. This also means supporting mania at their peak when the arguments starting to bring them down are flinching from reality. I only have brief experience with bipolar, because the only person I know who had bipolar disorder stopped having it after one cycle of paying attention to call her out on her nonsense at the relevant phases. She's quite happy and stable now, without mood stabilizers, so it looks like at least in her case it really was as simple as that.

Noisy hardware

Let's try a harder one. What about schizophrenia?

I'm not an expert on schizophrenia. I haven't even met anyone suffering from it. I've just skimmed through some papers on how it looks from a predictive coding perspective. Normally, this is where I'd say "Therefore take my perspective with a grain of salt", except that my whole thesis is "Don't do things you can't actually expect to be worth the risk, after taking into account

whether your own perspective is worth anything and whether it's safe to even consider whether your perspective is worth anything". So instead of telling you to take what I say with a grain of salt, I'm just going to double down: don't do things you can't actually expect to be worth the risk, after taking into account whether your own perspective is worth anything and whether it's safe to even consider whether your perspective is worth anything.

With that said, from my relatively uneducated perspective on schizophrenia, it *looks* like "Confidently predicting that everything is unpredictable (at low levels), and maintaining confidence at higher levels of abstraction through insecurity about everything falling apart". As in, "Of course they're closed to the idea that the CIA isn't messing with them. The alternative is to believe that their brain is failing to make sense of things. And if they let that in they might fail to make sense of anything -- because if they can't expect their brain to work then that opens up uncertainty faster than they can resolve it. And because if they can't expect their brain to work they can't intend for it to work, and if they can't even *try* to make things make sense then they're kinda SOL. They may have delusional beliefs, but at least they are coherent enough to be modeled as *having* beliefs".

Even if this isn't what schizophrenia is, this is the kind of thing you'd expect to happen if you had a brain that's just unable to make sense of its incoming data due to noise in the data -- perhaps including processing noise coming in at higher levels of processing. If your brain physically isn't working, you're going to have a hard time without a physical fix. But still, "The CIA is after me" is a belief, we disagree with it, and the difficulty of getting that to respond to evidence sure looks like a security limitation.

Recognizing that, and turning towards the uncertainty of how necessary that flinch is -- and how to make it less necessary -- provides a path to explore for anyone looking to help someone with this kind of problem find enough security to engage with the reality of their situation -- and therefore a more effective way of dealing with this likely shoddy hardware.

What would it be like *for you* if you were observing things that looked like the CIA were after you, and also like you just can't trust your own brain because you're losing your mind? "What to do if you're losing your mind?" is a terrifying question to ask, because you need your mind to answer it, and you can't trust *that*. Put yourself in those shoes, figure out how *you* could handle being in that situation, and you might have a path to helping the schizophrenic deal more effectively with their genuine hardware struggles.

Or maybe not. Again, this part is empirical, and on you. The point is just that once again, even when hardware is malfunctioning and making sense-making hard, there will still be disagreements to navigate -- just *more* of them, and with more difficulty of navigation.

Simple overload

Hardware doesn't have to be banged on or diseased in order to fail to perform. The same way a noncripple will kinda act like one after forced to run to the point of exhaustion, our brains can only do so much. We all have our limitations.

One of the more interesting experiences I've had in the last few years is of losing my mind. Partially.

I was dealing with a complex, high-stakes situation involving things very dear to me that required my full attention and processing capacity for over a month straight. I had to basically learn a new field, find results that the experts couldn't find, and then manage to earn the security/respect/attention needed to get them to accept my answers despite not having time to hash out our disagreements and explain how I knew my answers to be correct. This is on top of more than my normal life responsibilities. This went on for over a month, every day, without breaks. By the end, *I* started to break.

Normally, we maintain a fairly good agreement between what we believe and what we believe we should believe. We may have an irrational fear here or there, or be not quite as kind as we tell ourselves we should be, but more or less things will track. We *make sure* they track, by updating when divergences become apparent. I recognize concepts like "irrational fear" to be nonsense and "I should be nicer" as not necessarily correct (and how to find the extent to which it is justified), so I'm used to having very good agreement between my meta-beliefs and object level beliefs. As time went on though, I could feel myself coming apart at those seams; I just couldn't repair the gap as fast as it was growing.

My parents are great. Not *perfect*, because no one is, but great. They were very helpful at the time and certainly doing their best and engaging honestly. The only thing I could think to do when they got something wrong was yell at them, which was obviously crazy. I simply didn't have the ability to compute a more sensible answer on the fly, and had to physically leave so that I didn't do something stupid trying to engage with more than I could manage; I stand by and endorse that act of insecurity. Okay, whatever, people get stressed sometimes and yelling -- or better yet, walking away so as to not yell -- isn't *that* crazy, even if it's totally out of place. It's at least understandable, so long as it's a one off thing.

It wasn't just that though.

I left to the garage to work on a knife I had been making for my mom, gluing the handles on in the vice. And walked into the already sharpened knife. Normally I'm not so stupid as to walk into knives (I promise), but that day I didn't have the capacity for normal caution. Normally I wouldn't be so comfortable with what had just happened, but at the time all I could think was

"Ooh! Human meat is red meat! Okay, where's the electrical tape". I had higher priorities. Places to be, things to do.

When driving to those places, I *knew* that I didn't have my wits about me enough to navigate there without error -- and it wasn't a complicated route, or a new one. I didn't even try to fix that, because whatever, driving safely is my priority and if I have to back track a bit after I miss a turn or two *that's fine*.

The scariest part, for me, was when I thought I was about to get in a fist fight, and couldn't do anything to stop it. Not scary because I was afraid of the fight, but because I was afraid of just how out of control I was of my own behavior.

Fist fights are fine as kids, but *way* not safe as full grown men, and absolutely not worth it. Normally I'm very good at deescalating by making it feel pathetic to lose control of oneself enough to take a swing in anger.⁵ This time, when this guy started yelling in my direction, I knew very well that I would not be doing that. I didn't want to fight. Fighting is crazy, and I was still very very aware of that. At the same time, I just knew that if it turned out that it was me he was yelling at and he pushed the issue, I was going to make an equally persuasive case that he was pathetic if he *doesn't* swing. More specifically, I was completely unwilling to let him take my already scarce cognitive resources and redirect them to helping him shield himself from the destabilizing idea that being a middle aged man running around yelling at people for no reason *is already pathetic* -- even if he were to stop and get his shit together. I was just going to share my honest belief that he's beneath my consideration, and let him to do what he will with that information. It's just that I knew it was unlikely he could actually disagree, and unlikely that he was secure enough to update on this information, so the response he was most likely to have to that destabilizing information would be to try to make it go away by physically forcing me to recant.

It's hard to describe what the experience was like. I couldn't sort out that internal conflict in real time, and couldn't just self censor like I normally could. It was like I was a mere observer in my own body, with confident and legitimately justified meta-beliefs about what the right thing to do was, and it just didn't matter because [that day I was not my elephant](#).

It's not that it "wasn't a disagreement", and it wasn't that it "couldn't be navigated" in some absolute sense. It took a month or so after it was all over, but I did return to baseline and would absolutely de-escalate that situation well today. So far as I can tell, every part of that is completely in line with the framework I offered here, and it's just a limit of how much information my brain could process.

At the time though, the result is that I just couldn't process more information. It's not that there wasn't time to say the words that point at the things, just that I couldn't process the emotional

updates. You could have been freaked out and trying to warn me of the danger in the path I was going down, and I would have told you -- with an emotionally blank face -- that I agreed 100%. You're right. And it doesn't matter, because there's no place for that information to go and nothing for it to connect to. Take a number.

Navigating with imperfect hardware

Take away sleep, nutrition, add processing noise, concussions, and everything gets harder. At some point I'd have to accept some losses in order to maintain sanity. The temptation was great, by the end of the month. When I finally did have to cut my residual losses it was the first time in a long time that I couldn't actually trust that I was making the object level best bet without being unduly swayed by temptation that I did not consider valid.

It's not that I couldn't update *any* beliefs, I just couldn't update them *all*.

I could have prepared for the unlikely event of some A-hole picking a fight with me, but it would have come at the cost of working on the more certain problems; I bet on it not being necessary. I could have let go of my anger towards arrogant idiots stepping out of line *on whole*, clearing a larger swath of potential problem, but that would have directly cost me the ability to get done what I needed done.

I could have learned something from walking into a knife, but again, it would have come at the cost of working on more important problems.

I did have a conversation with my parents, because it was important to keep that relationship working well in the short term too, but it took setting aside some time. I couldn't set aside enough time to cohere my beliefs enough to do it *well*, so I had to add a meta note of "I'm sorry I still sound kinda angry, I don't actually blame you I'm just legit struggling to hold it all together".

I had to prioritize. I had to track where I was flinching, which of my beliefs were wrong, and which disagreements were worth resolving when. This meant maintaining accurate meta-beliefs about my own inaccurate beliefs and their likely impacts, so that I could make reasoned choices with a brain that was physically incapable of updating on all the important information available to me -- and was gradually going crazy.

Thankfully it worked out.

I was able to mitigate the costs well and my bets paid off. I didn't get into any car crashes, or spurious fist fights, or damage my relationship with my parents. I do have a new scar, but it's small.

I got the results I needed to get, and I recovered my sanity. It was worth it.

Physiological limitations are real. Some of them influence the degree to which our brains can work as inference engines, satisfying the requirements for successfully navigating disagreements by "just showing relevant evidence". When we can't meet those requirements and update on evidence, then we just can't -- at least, without physically repairing our brains. Such is life, sometimes.

At the same time, even in those cases, knowing how to navigate matters. Sometimes it matters more.

Notes

1. While it has beliefs about how much power is "too much", it doesn't have any beliefs either way about whether this object level belief is *correct* -- and no place to put such beliefs and no ability to form them. ↩
2. At a cost, remember. There's always a reason for things, and just because I chose to stop depressing there doesn't mean there wasn't anything to resolve, or that I was able to resolve it on the spot. Just that I decided on the spot that it wasn't worth trying to resolve at the time.

Applied to the nutritional-deficiency-induced-depression case, this would look like "let go, and be content to sit around with no energy -- until I can make it to a nutritionist" .

↩

3. I know some people with similar experiences simply deciding to lose 70lb or quit smoking packs per day -- once they had a strong enough reason to do so. ↩
4. Similarly, [agentyduck](#) explicitly works through the same thing for social anxiety. I'm particularly amused by this part:

"During the conversation in which I described my plan to him, we meandered to the topic of a meetup of professional hypnotists he'd recently attended. He told me they talked in passing about what it's like to change their own behaviors. They all knew they could use a long, draw-out induction (or series of inductions and post-hypnotic suggestions) to self-modify if they wanted. But that takes time and energy, and it turns out that if you're sufficiently confident it'll work... you don't have to bother with the hypnosis.

Think about that for a minute. They treated it as a perfectly normal, every-day occurrence. Basically they were saying, "Yeah, when I don't like what System 1 is doing, I just tell it to do something else instead. No biggy." They seem to have this available as a primitive action."

Because yes, that's exactly what it's like. And you don't even have to be a hypnotist to do it. ↩

5. Implicitly communicating "You're better than that, I know you are. You know you are too."

Even if you wouldn't have been without the nudge, it's hard to counter "No I'm not!". It's also hard to counter "Losing my temper is totally a grown man thing to do!". Especially when you know the other person is still going to be looking at you like he sees right through you. ↩

On failure, and keeping doors open; closing thoughts

A lot of what motivated the approach I've taken in [this sequence](#) has been a desire to avoid predictable failures - to find things to do that I don't know to be dumb, relative to what I know and care about. When most of the "technique" is just "don't try to do dumb things," it makes "failure" conceptually weird. How do you measure success if it comes primarily by ditching stupid goals?

When success comes from picking winning battles, 'failure' gets harder to define

For example, if you're not careful you might think "Neat story about the [irrational fear of heights](#), but it's anecdotal. Maybe cherry picked. Let's see an RCT to see what fraction of the time you succeed in fixing irrational fears!". The problem with this line of thinking is that the fraction is necessarily 0/0. I couldn't *possibly* have failed to fix her irrational fear for the same reason I couldn't possibly have succeeding in fixing it; you can't fix what ain't is. The whole point of that story is that she *never had* an irrational fear, and it was the illusion that she did which preserved that internal conflict in the first place.

The [fire poker example](#) is an even clearer example. The kid literally had no problem with the pain, but because no one there could notice this, they inadvertently pressured him into not noticing either. Again, the only "technique" was not getting fooled by the nonsense problem statement, and therefore not trying to do something foolish like "help him not feel pain".

A reasonable response to noticing this is "Aha! Let's instead measure how effective these insights are at resolving disagreements!" -- which is a good idea. But the concept applies here as well, complicating things further.

In [the pool situation](#) I wasn't trying to "resolve the disagreement", I was just playing -- and it's exactly that playful lack of attachment to any outcome that enabled the resolution of the disagreement. If I had instead tried to "resolve the disagreement", then what happens to any hint that she's not interested in resolving the disagreement?¹ Or that *I* am not interested in doing what it'd take, for that matter? Either I take it into account and *stop trying to resolve the disagreement*, or else I shut out information about why I might fail -- which is exactly the kind of information that is important in order to prevent failure. If I were to ignore that information, then any push in that direction requires pushing back with "No, you *need* to provide a convincing rebuttal, or else change your mind", and that isn't the attitude that tends to convince people they want to talk to you.

But still, it seems like there's something to the question.

What failure is left

For example, if I were to tell you I fainted last time I got my blood drawn, it'd seem like "Wtf? I can buy being afraid of heights until you check your knots, but I have a hard time buying that you thought that was a good time for a nap". It really does seem pretty simple with an obvious right answer, which means there shouldn't be any disagreement or fainting going on. It's not that this *can't* happen based on what I've said so far, but that it really shouldn't be a very stable sort of affair, so if it manages to show up *at all* there's something to be explained, and it seems to deflate the apparent usefulness of this sequence.

If, on the other hand, I told you that I *used to* faint when I got my blood drawn, and then over the course of figuring this stuff out, I stopped flinching against the reality that I *had* this response, and that once I stopped flinching I found it pretty funny and I stopped having that vasovagal response... that'd seem like more what you'd expect if what I've been talking about is real and useful.

If I were to say that *both* were true -- that I used to faint, that it went away the moment I looked at it for what it was *and* that it came back and I fainted again last time -- then it'd be cherry picking to talk about one over the other. So the concept has merit.

In reality neither of these are quite true, but something like both of them is. Since this post is about failure, let's pick that one to talk about. This is an example of something that happened very recently, after writing the rest of this sequence. I decided that rather than give you a polished conclusion, I'd work through this failure *with* you, so you can see a bit of what the thought process is like when things *aren't* going well, and how to make sense of things when that happens.

Vasovagal struggles

What *should have* happened, is nothing. Or rather, nothing of note. In *reality*, I didn't actually faint, and it wasn't about getting blood drawn, but I *was* recently surprised by having an overactive vasovagal response in a new context. Enough that I had to stop and address it, and that I wasn't sure I *wouldn't* faint.

I felt sufficiently sure that it wasn't an appropriate response that I should be able to "just not do it" and would normally expect to be able to demonstrate trust in myself in situations like that. It's not like "it was more blood than I'm used to" or anything -- I've handled worse situations completely without issue, so there's something else that was going on. Before figuring out what

is actually going on, it sure *looks* like something that shouldn't happen, and is therefore evidence either that I'm failing to follow my own framework or else that the framework is incomplete -- either way, somewhat damning for the usefulness of reading this sequence.

The way the experience went was something like this:

Oh great, vasovagal response; this isn't helpful. Okay, let's make sure we're not freaking out over nothing, and compose ourselves. Okay, yeah, that's not helping. I'm unable to "just decide" to do something different and have my body follow, which means I'm out of touch with why I'm doing what I'm doing. Why am I choosing to do this?

I was willing to listen, but I had no answer. Not even an opaque [felt sense](#) that could be weighed or focused on. This the exact thing that shouldn't happen if I ~~practice~~ grok what I preach. *And* it was happening.

It had been a while since I had an experience like that. Usually, the sense of "My brain is doing the wrong thing" *does* very rapidly resolve into either choosing to do something different, or else a recognition that I actually don't know what the right thing to do is and a focus on figuring out what to do on the object level.

But in this case I really was stumped. My own behavior seemed obviously dumb, *and* it wasn't immediately resolving when I noticed that *and* the lack of immediate resolution wasn't immediately convincing me that *maybe it's reasonable after all*. I was genuinely interested in why I was doing it, but I wasn't at all expecting to be convinced which is the cue to be the one speaking, and that just wasn't working. I was genuinely failing to relate to the experience as "just unsure what to do", and therefore unable to live up to the aim of "not having psychological problems, just real ones".

This was a little annoying. Especially just having written a sequence on why it basically shouldn't happen once you understand where "psychological problems" come from.

It's not that "Oh no, I have a disagreement I wasn't able to resolve"; that happens. But even with things like allergies, when I think "This is dumb, I shouldn't do allergies", I actually manage to stop being congested... only to have an immediate sneezing fit and realize that I prefer being congested, but still. There's at least an unbroken thread of "I can actually change this if I know what I want to do instead", and last I left it was at "Actually... how *do* I know whether this is a response to something genuinely benign like pollen in the air, or whether I'm fighting off a cold?" -- which is exactly the kind of thing that *should* happen, when it turns out that changing behavior isn't as easy as it seems like it should be.² It's supposed to be about not knowing what to do, not about not knowing how to do it.

So afterwards I kept wondering. *What is the purpose of the vasovagal response?* I may not have been able to experience it as "being unsure what to do" in the moment, but I still recognized it as a disagreement between parts of my own brain, even if I wasn't able to functionally corral it into one coherent "I" in the moment -- so I was *curious*.

I don't buy the "[to play dead from predators](#)" just so stories. It feels much more like a symptom of poor regulation than anything intentional, but there still must be a purpose for *that direction*, or else there'd be no reason to ever err in that direction. So what's *that* about?

Another hypothesis I found was that it's compensation for sympathetic nervous system activation. "It's an overcompensation" -- but that doesn't feel right either. My heart rate didn't spike and then tank. There *never was* that sympathetic nervous system activation needing compensation, so that doesn't fit.

But I had a hint.

Back when I had the initial bit of vasovagal response due to getting needles stuck in me, one of the nurses said "It's something that actually happens more often to 'tough guys' because they're trying to over-control their responses". It's a feel good reassurance thing to tell someone so I didn't actually believe it, but it stuck with me because it was interestingly plausible. So I looked up what it correlates with. Motion sickness.³ Sensory sensitivity. Okay, that's me. Neat.

Okay, so "over-controlling" fits. Perhaps it's not that I'm "trying to almost faint", and not "responding too strongly to sympathetic nervous system activation" but rather that I'm *avoiding* sympathetic nervous system activation.

...That fits.

The original sensitizing event makes a *lot* more sense now.

The first time I ever had this response came as I watched a nurse try and fail to set up an IV when I was really sick with pneumonia. I watched her dig around with the needle for thirty seconds or so (Because why not watch? It's just information, right?), then pull the needle out, and stick it in to dig around for another minute or so. I thought it was "just too much". I don't think that's it anymore.

So far as I can tell now, it's specifically that she was *incompetent*, and I would have been better off taking the needle and doing it myself. I have very large veins in my arms, to the point where even when dehydrated the big one is still impossible to miss. I pointed this out to her, and she told me "That's where we take blood out. We put IVs in at the wrist". When she finally left, she muttered something about it not being her lucky day on her way out. She must have said the

same thing to the nurse she got to replace her, because that nurse came in saying "Lucky day, my ass! Look at this huge vein! Let's use this one."

The suppressed sympathetic activation had become clear.

Get the fuck out of here and let me do it, you incompetent fool

Testing it against other situations where I had no issue, it fits. I could put myself back in that initial situation and feel the vasovagal response. If I swap in someone I personally know to not be an idiot, and competent with a needle, it's not a concern. In all the situations that were "worse" where I wasn't squeamish about a pint of blood or whatever, *I was in control* -- or at least, doing my best to be, whether I succeeded or not. I had things to do which *justified* a sympathetic nervous system response -- even if that thing to do was nothing more than the equivalent of "Watching to see whether it's time to tell the nurse to get another nurse who knows how to put an IV into an obvious and large vein".

What went wrong

With that purpose in mind, I could feel what I could do. Putting myself back "in the driver seat", even just sitting there thinking about it, resulted in perceptibly higher blood pressure in my face. Part of that was that the mindset came with more core tension, but even when I relaxed my core I could still feel it. That action of "let's compose ourselves here, and make sure we're not freaking out unnecessarily" *was* the blood pressure lowering move; I was fighting against the response I *would have had*, if I were just to drop all inhibitions. I wasn't able to "choose not to" because I had my signs wrong. I wanted to fix problems caused by pushing the stick too far down, so I pushed the stick... down. Like an idiot.

So why would I do that? What was I thinking? What bad would happen if I *didn't* suppress the sympathetic nervous system response?

Oh.

Right.

That.

In the [previous post](#) I talked about gradually losing my mind over a particularly stressful month, and finally getting to the point of feeling tempted to give up and cede control without endorsing that as a good decision. Increasing my blood pressure would have meant going back to *that*. Going back to trying to take control of a situation that I would struggle to stay on top of even at

max capacity, which I had become pretty averse to by the end of the last one -- and in a case where that probably wasn't necessary given what I knew at the time. And in hindsight turned out to not be necessary.

That was the flinch. I hadn't gotten around to going back and sorting out when it *does* make sense to decline to take on that stress, and what it takes to justify it, so I was defaulting to not doing it and wasn't sufficiently open to *considering* doing it. So I couldn't see my own stupidity, and couldn't avoid this failure mode.

Okay, so if that's what happened, then what?

How is it that I ended up in a situation where I wasn't able to relate to the problem as an object level uncertainty, despite knowing that it's fundamentally a disagreement and having worked out what my options are and where to look for resolution?

Does it point at something missing or invalid in my framework? At something in my own framework that I failed to grok?

I'm still trying to work this out for myself, as I write this.

It's obviously an example of [the perils of navigating blind to maps](#), due to the whole "Oh no, BP dropping, push the stick down further" idiocy. But also, more importantly, that error was enabled by the fact that I was more focused on getting *away* from an undesired response than I was at getting towards anything. It was ultimately caused by an insecurity noticing "If I don't flinch from this opportunity I might engage too much and lose too much sanity" -- and I didn't recognize the insecurity flinch for a while.

Why didn't I notice the flinch?

That's a good question. I don't really know.

I guess I didn't recognize that I was aiming *away* from a certain response rather than aiming *for* anything in particular -- or maybe, didn't notice the importance of this because I didn't notice "Don't wanna deal with this shit, and I don't think I have to or should" as insecurity. Or, as a *problematic* insecurity, I guess? It pretty clearly is a flinching from the firehose of information, but it's for a different reason. Instead of "I'm afraid of what I might find" and "I have strong reason to find out", it's "just too fatiguing" and "I really don't have sufficient reason to". Or.. is that true though? Since I *did* have reason to -- I just didn't believe it was valid.

That fits. I was aware that there was an imperfection in my map there, I just didn't think that it *mattered* nearly as much as it did, so I didn't even think of it as an "insecurity" even though it was.⁴

I guess the way I'd say it is that our own minds are part of the territory, and there's legitimate uncertainty about what's going on in there too. Knowing the structure helps you throw out limiting (and *false*) beliefs, and helps you notice where to look. And that *does* lead to a lot of things -- *most* things, even -- feeling a lot more like "I'm not sure what to do" than "I gotta figure out how to get this brain to do the right thing". But it also takes skill, experience, and clearing out mistaken beliefs in order to get there. So even with practice, there'll be occasional failures -- for the same reason you can still find the occasional bug even in relatively mature software. They'll be fewer and further between, but there are always more unexplored corner cases to bump into.

I think I'm happy with that.

I did fall off, but I think I'm back on track.⁵

And I don't see any way to have gotten back on track more quickly this time, though I see ways to do it faster next time.

When I originally said "You'll never struggle with irrational fears!" in the intro, I was kinda laughing about how it's a little bit cheating because it's not like you get to magically stop feeling fear in the face of real danger that you haven't taken seriously yet, or that you get to figure out all disagreements let alone easily.

At the same time, a lot of formerly "sticky" problems really do dissolve.

And even when you do genuinely fail at something you took your best shot at, you'll still have a path available to figure out why you failed. In case it's important enough to figure it out, should it come up again.

Notes

1. Like when she responds with "Are you calling me ugly?" rather than immediately addressing the inconsistencies between her beliefs that I had highlighted. ↪
2. A few days after writing this I had what I thought was allergies, decided to be a bit less congested and managed that without a sneezing fit -- and it turned out to be a cold. ↪
3. Motion sickness is another one I've struggled with which seems "obviously wrong" enough that it should be a trivial fix, and hasn't been. It's not a big enough issue that I've ever sat down when *not* motion sick and tried to sort things out, but in the moment I've only had partial success and more failure than success in matching what seems like the right outcome (i.e. not feeling sick) with my actual behavior (i.e. feeling sick). It's kinda by definition when I'm already overwhelmed, and more often when low on sleep, so it's not really *that* surprising, I guess. Perhaps I should sit down and address it just out of curiosity to see what happens. ↪

4. I did mention that insecurities [can be very subtle](#). ↩

5. Since writing this, I have had a bit of a test. Not quite the same thing and perhaps less triggering so who knows, but it went well. ↩