Hammertime

Hammers and Nails

This is a linkpost for https://radimentary.wordpress.com/2018/01/22/hammer-and-nails/#more-2328

If all you have is a hammer, everything looks like a nail.

The most important idea I've blogged about so far is Taking Ideas Seriously, which is itself a generalization of Zvi's More Dakka. This post is an elaboration of how to fully integrate a new idea.

I draw a dichotomy between Hammers and Nails:

A *Hammer* is someone who picks one strategy and uses it to solve as many problems as possible.

A *Nail* is someone who picks one problem and tries all the strategies until it gets solved.

Human beings are generally Nails, fixating on one specific problem at a time and throwing their entire toolkit at it. A Nail gets good at solving important problems slowly and laboriously but can fail to recognize the power and generality of his tools.

Sometimes it's better to be a Hammer. Great advice is always a hammer: an organizing principle that works across many domains. To get the most mileage out of a single hammer, don't stop at using it to tackle your current pet problem. Use it everywhere. Ideas don't get worn down from use.

Regardless of which you are at a given moment, be systematic because Choices are Bad.

Only a Few Tricks

I am reminded of a classic speech of the mathematician Gian-Carlo Rota. His fifth point is to be a Hammer (emphasis mine):

> A long time ago an older and well known number theorist made some disparaging remarks about Paul Erdos' work. You admire contributions to mathematics as much as I do, and I felt annoyed when the older mathematician flatly and definitively stated that all of Erdos' work could be reduced to a few tricks which Erdos repeatedly relied on in his proofs. **What the number theorist did not realize is that other mathematicians, even the very best, also rely on a few tricks which they use over and over.** Take Hilbert. The second volume of Hilbert's collected papers contains Hilbert's papers in invariant theory. I have made a point of reading some of these papers with care. It is sad to note that some of Hilbert's beautiful results have been completely forgotten. But on reading the proofs of Hilbert's striking and deep theorems in invariant theory, it was surprising to verify that Hilbert's proofs relied on the same few tricks. Even Hilbert had only a few tricks!

The greatest mathematicians of all time created vast swathes of their work by applying a single precious technique to every problem they could find. My favorite

book of mathematics is The Probabilistic Method, by Alon and Spencer. It never ceases to amaze me that this same method applies to:

1. (The Erdős–Kac Theorem) The number of distinct prime factors of a random integer between 1 and n behaves like a normal distribution with mean and variance $\log \log n$.
2. (Heilbronn's Triangle Problem) What is the maximum $\Delta(n)$ for which there exist n points in the unit square, no three of which form a triangle with area less than $\Delta(n)$?
3. (The Erdős-Rényi Phase Transition) A typical random graph where each edge exists with probability $\frac{1}{n^c}$ has connected components of size $O(\log n)$. A typical random graph where each edge exists with probability $\frac{1}{n^c}$ has a giant component of size linear in n.

It's amusing to note that in the same speech, Rota expounded the benefits of being a Nail just two points later:

> Richard Feynman was fond of giving the following advice on how to be a genius. You have to keep a dozen of your favorite problems constantly present in your mind, although by and large they will lay in a dormant state. Every time you hear or read a new trick or a new result, test it against each of your twelve problems to see whether it helps. Every once in a while there will be a hit, and people will say: "*How did he do it? He must be a genius!*"

Both mindsets are vital.

To be a Nail is to study a single problem from every angle. It is often the case that each technique sheds light on only one side of the problem, and by circumambulating it via the application of many hammers at once, one corners the problem in a deep way. This remains true well past a problem's resolution - insight can continue to be drawn from it as other methods are applied and more satisfying proofs attained.

Usually even the failure of certain techniques sheds light on shape of the difficulty. One classic example of an enlightening failure is the consistent overcounting (by exactly a factor of two!) of primes by sieve methods. This failure is so serious and unfixable that it has its own name: the Parity Problem.

Dually, to be a Hammer is to study a single *technique* from every angle. In the case of the probabilistic method, a breadth of cheap applications were found immediately by simply systematically studying uniform random constructions. However, particularly adept Hammers like Erdős upgraded the basic method into a superweapon by steadfastly applying it to harder and harder problems. Variations of the Probabilistic Method like the Lovász Local Lemma, Shearer's entropy lemma, and the Azuma-Hoeffding inequality are now canon due to the persistence of Hammers.

Be Systematic

The upshot is not that Hammers are better than Nails. Rather, there is a place for both Hammers and Nails, and in particular both mindsets are far superior to the wishy-washy blind meandering that characterizes overwhelmed novices. There may be an endless supply of advice - even great advice - on the internet, and yet any given person should organize their life around systematically applying a few tricks or solving a few problems.

Taking an idea seriously is difficult and expensive. You'll have to tear down competing mental real estate and build a whole new palace for it. You'll have to field test it all over the place without getting superstitious. You'll have to gently titrate for the amount you need until you have enough Dakka.

Therefore, be a Hammer and make that idea pay rent. Hell, you're the president, the emperor, the king. There's no rent control in your head! Get that idea for all its got.

Exercise for the reader: all things have their accustomed uses. Give me ten unaccustomed uses of your favorite instrumental rationality technique! (Bonus points for demonstrating intent to kill.)

Hammertime Day 1: Bug Hunt

Rationality is systematized winning.

In Hammers and Nails, I suggested that rationalists need to be more systematic in the practice of our craft. In this post, I will use the word Hammer for a single technique well-practiced and broadly applied.

Hammertime is a 30-day instrumental rationality sequence I am designing for myself to build competence with techniques. Its objective is to turn rationalists into systematic rationalists. By the end of this sequence, I hope to upgrade each Hammer from Bronze Mace to Vorpal Dragonscale Sledgehammer of the Whale. I invite you to join me on this journey.

The core concept: One Day, One Hammer.

Hammertime Schedule

In Hammertime, we will practice 10 Hammers over 30 days. Each exercise is scalable from a half hour to an entire day. The Hammers will be bootleg CFAR techniques:

1. Bug Hunt
2. Resolve Cycles
3. TAPs
4. Design
5. CoZE
6. Mantras
7. Goal Factoring
8. Focusing
9. Internal Double Crux
10. Planning

There will be three cycles of 10 days each, practicing each technique a total of three times. The first cycle will cover basics and solve bugs at the life-hack level. The second cycle will reinforce the technique, cover variations and generalizations, and solve tougher challenges. The third cycle will build fluid compound movements out of multiple core techniques.

Day 1: Bug Hunt

A bug is anything in life that needs improvement. Even if something is going well, if you can imagine it going better, there's a bug.

On the first day of Hammertime, we will scour our lives with a fine-toothed comb to find as many bugs as possible. A comprehensive bug list will provide the raw material on which we practice every other rationality technique. For the first cycle of Bug Hunt, look for small, concrete bugs. The whole exercise should take a bit over an hour.

WARNINGS: Focus on finding bugs, not solving them. If you can solve the bug immediately, go for it. Otherwise, hold off on proposing solutions. Writing down a bug

does not mean you commit to doing anything about it.

1. Setup

Find a notebook, phone app, spreadsheet, or Google Doc to record your bugs - preferably something you can bring with you throughout the day. We will refer back to it repeatedly in the coming days for bugs to solve.

During Bug Hunt, spend the next 30 minutes writing down as many bugs as you can. Following each of the six sets of prompts in the next section, set a timer for 5 minutes and list as many bugs as you notice.

2. Prompts

A. Mindful Walkthrough

Walk through your daily routine in your head and look for places that need improvement. Do you get up on time? Do you have a morning routine? Do you waste mental effort deciding whether to or what to eat for breakfast? Do you take the most efficient commute, and make the most of time in transit?

Fast forward to work or school. Are there physical discomforts? Are you missing any tools? Are there particular people who bother you, or to whom you don't speak enough? Do you ask for help when you need it? Do you know how to shut up? Is there unproductive dead time during meetings, classes, or builds? Do you take care of yourself during the day?

Think about the evening at home. Do you waste time deciding where or what to eat? Are there hobbies you want to try? Are there things you know will be more fun that you're not doing? Do you progress consistently on your side projects? Do you sleep on time? How is your sleep quality?

B. Hobbies, Habits, and Skills

Walk through the things you do on a regular basis. Are there habits you mean to drop? Are there habits you mean to pick up but never seem to get around to?

For each hobby or habit, answer the following questions. Do you do it enough? Do you do it too much? Are there ways you could improve your experience? Do it in a different place and time? Do it with other people or alone?

Perhaps you have skills to practice. Are you as good as you want to be? Do you practice regularly? Have you plateaued by overtraining? Are there minor recurring discomforts keeping you from trying? Are there directions you haven't tried which might indirectly improve your abilities?

C. Space

Look around your living space, your workspace, or the interior of your vehicle. What would you change?

Space should be functional. Is there clutter you circumnavigate on a daily basis? Are your chairs and tables at the right height? Is your bed comfortable? Are there towels, pans, notebooks, or papers sitting out taunting you? Are there important things that deserve a more central position? Have you set up Schelling places for glasses, wallets, and phones?

Space should be aesthetically pleasing. Do pieces of furniture or equipment stick out comically? Do your walls feel drab and depressing? Are there carpet stains or dust mites that keep catching your eye and sucking out your happiness? Are you tired of the art on the walls?

Space on the monitor can be as important as physical space. Do you have enough screens? Do you find yourself repeating mechanical boot-up and shutdown sequences that can be automated? Do you use all the browser extensions and keyboard shortcuts? Is there a voice in the back of your head whispering at you to learn vim?

D. Time and Attention

People and things clamor for your attention. What's missing from your life that would let you live as intentionally as possible?

Many activities are bottomless time sinks. Do you watch shows or play games you no longer enjoy? Do you get dragged into conversations that hold no value? Do you find yourself rolling the mouse wheel down endless Facebook or Reddit feeds? Are there classes, meetings, commutes, or projects that zombify you for the rest of the day? Do you set up ejector seats in advance to protect yourself from time sinks?

Focus on the things you don't pay enough attention to. Do you often make mistakes on autopilot? Are there friends or family you've neglected or grown distant from? Are there conversations you zone out in that you could get more out of? Is there a childhood dream you've forgotten?

Sometimes trivial distractions lead to spectacular failures. Are there slight, recurring physical discomforts that drain your agency? Does the temperature outside prevent you from exercising? Is there something shiny that always draws your eye away from work?

E. Blind spots

Our biggest bugs can hide in cognitive blind spots.

Outside view your life. Are you sufficiently awesome? What is your biggest weakness? If there is one thing holding you back from achieving your goals, what would it be? Do you have mysterious attachments to pieces of your identity? Do you routinely over- or under-estimate your own ability?

Simulate your best friend in your head. What do they say about you that surprises you? What behaviors annoy them? What behaviors would they appreciate? Is there a piece of advice they keep giving you?

Summon your Dumbledore. What would he say to you? What deep wisdom are you blind to? If you were the protagonist, what genre would this life be?

Look to admiration and jealousy for insight. Are you the person you most admire? What skills and traits do others have that you want?

F. Fear and Trembling

The shadows we flinch away from can hide the most bountiful treasures.

What are your greatest fears and anxieties? Do you have the strength to be vulnerable? Are there necessary and proper actions you need to take? Are there truths you're scared to say out loud? What do you lie to yourself about?

Look to your social circle. Are there good people you hide from? Are there conversation topics that cause you scramble away? What do people say that cause you to lose your composure?

Look to the past and future as far as your eyes allow. What deadlines cause you to avert your eyes? Is there a kind of person you are terrified of becoming? Or are you most afraid of stagnation? Do you trust your past and future selves?

3. Sort

Hopefully, you came up with at least 100 bugs; I came up with 142. Time for some housekeeping. Input your bugs into a spreadsheet to organize and coalesce similar ones. Using System 1, assign difficulty ratings from 1 to 10, where 1 is "I could solve it right now" and 10 is "Just thinking about it causes existential panic." Sort them in increasing order of difficulty.

In the coming days, we will go down the list systematically, hitting as many nails as possible with each hammer.

Daily Challenge

To help others brainstorm, share your strangest bug-fix story. I'll start:

The muscles on the left half of my face are more responsive, which caused me to smile asymmetrically for most of my life. Therefore, my usual smile wasn't far from a contemptuous smirk, and caused me to feel dismissive of everyone I smiled at. I trained myself to smile on both sides and now feel warmer towards people.

Hammertime Day 2: Yoda Timers

This is part 2 of 30 in the Hammertime Sequence. Click here for the intro.

> *No! Try not! Do, or do not. There is no try.*
> —Yoda

There's a copy of Barney Stinson in my head who pops up every so often to say: "Challenged Accepted!" When Eliezer wrote about the biggest mistakes in the Sequences, my inner Barney started bouncing off the walls. Hammertime is a sequence designed to correct the three top mistakes by:

1. Creating a program to actually *practice* rationality.
2. Emphasizing doing better in everyday life.
3. Focusing on rational action instead of rational belief.

This is going to be legen ... wait for it ... dary!

Day 2: Yoda Timers

> *Look, you don't understand human nature. People wouldn't try for five minutes before giving up if the fate of humanity were at stake.*
> —Use the Try Harder, Luke

The Yoda Timer (CFAR calls it a Resolve Cycle) has three simple steps:

1. Pick a bug.
2. Set a timer for 5 minutes.
3. Solve the bug.

Motivation

Before we begin, I want to call attention to two ways to make the most of Yoda Timers.

1. The One Inch Punch

Pick something you're afraid of doing. Suppose I told you, "Try!" Try as hard as you can. What does that feel like?

Now suppose I told you, "Just do it!" Actually go and get it done. What does that feel like?

To me, trying feels like pushing hard against my own resistance. Doing feels like pushing hard *against reality*. The Yoda Timer is designed to teach (or remind) you to notice what pushing against reality feels like.

Bruce Lee was famous for his One Inch Punch, which had such explosive power because every muscle in his body aligned into the punch:

> The *one-inch punch* is a skill which uses *fa jin* (translated as explosive power) to generate tremendous amounts of impact force at extremely close distances. This "burst" effect had been common in *Neijia* forms. When performing this one-inch punch the practitioner stands with his fist very close to the target (the distance depends on the skill of the practitioner, usually from 0–6 inches, or 0-15 centimetres). Multiple abdominal muscles contribute to the punching power while being imperceptible to the attacker. It is a common misconception that "one-inch punches" utilize a snapping of the wrist. The target in such demonstrations vary, sometimes it is a fellow practitioner holding a phone book on the chest, sometimes wooden boards can be broken.

When you're in doing mode instead of trying mode, the inner conflicts fall away and you can practice punching reality with your whole soul. Imagine how far you'll go if every move you make carries the entire weight of your being.

2. Lateral Thinking

It's easy to get tunnel vision and freeze up with only 5 minutes to go. To get maximal effect out of Yoda Timers, however, you'll need to get more creative, not less. If you had to fix the bug in five minutes to save the world, what rules might you break?

To get you started, here are a few classic approaches: How much money will make the problem go away? What email or phone call could you make? What external reward, punishment, or commitment can you set up in five minutes that will guarantee the thing gets done? What alternative course of action would achieve the same desired effect?

3. Permission to Try

If there's something you can do in five minutes to improve your life, as a fellow human being I grant you permission to do it.

Five for Five

Pick your 5 easiest bugs from yesterday's Bug List.

WARNING: There only one valid reason to skip a bug - if you're uncertain whether you actually want to fix it. Later on, we will practice techniques for resolving inner conflicts. Difficulty is not a valid excuse to skip.

For each one, set a Yoda timer for five minutes and do it. That's it. Just do it.

If it helps, imagine that Yoda is watching. Yoda doesn't care how hard you try.

Daily Challenge

Share your most successful Yoda Timer bug-fixes.

Here are seven things I did in the past couple days with Yoda Timers:

1. Move furniture around and store unused junk to double effective floor space.

2. Train myself to place my glasses on one fixed countertop in the apartment.
3. Practice keyboard shortcuts for archiving emails and Chrome tab management.
4. Send all the emails and messages I plan to (and already can) send in the next week.
5. Order a white noise machine on Amazon and open the blinds to let in sunlight in the morning to optimize sleep schedule.
6. Practice keeping a pen in my pocket at all times so I can penspin instead of picking my face.
7. Plan and outline Hammertime.

Hammertime Day 3: TAPs

This is part 3 of 30 in the Hammertime Sequence. Click here for the intro.

A running theme of Hammertime, especially for the next two days, is *intentionality*, or deliberateness. Instrumental rationality is designed to inject intentionality into all aspects of your life. Here's how the 10 techniques fit into the intentionality puzzle:

1. Noticing and having more intentions (Bug Hunt, CoZE, TAPs).
2. Resolving internal conflict about what you intend to do (Goal Factoring, Focusing, Internal Double Crux).
3. Learning how to convert intention to action (Yoda Timers, TAPs, Planning).
4. Injecting intentions into System 1 so you can do what you intend even when you're not paying attention (TAPs, Design, Mantras).
5. Injecting intentions into reality so that reality pushes you towards, and not away from your goals (Design).

Trigger-Action Plans (TAPs) are the if-then statements of the brain. Installing a single TAP properly will convert a *single* intention into *repeated* action.

Day 3: TAPs

Recommended background reading: Making intentions concrete – Trigger-Action Planning.

1. TAPs 101

TAPs are micro-habits. Here's the basic setup:

1. *Pick a bug.* Again, skip bugs you're conflicted about.
2. *Identify a trigger.* An ideal trigger is concrete and sensory, like "water hitting my face in the shower," or "when I press the elevator button."
3. *Decide on an action you want to happen after the trigger*. Pick the minimum conceivable action that counts as progress towards solving the bug. Thus, "look at the stairwell" is better than "go up the stairs," and "sit up in bed" is better than "force myself out of bed."
4. *Rehearse the causal link.* Go to the trigger and act out the TAP ten times. If the trigger is not currently available, visualize it. Focus on noticing and remembering sensory data that will help you notice the trigger.
5. *Check the TAP in a week.* Write down the TAP when you intend to do it, and check back in a week to see if its installed. TAPs can require a lot of tweaking.

TAPs take a couple days to install successfully. Today, we will practice installing two TAPs.

2. The Sapience Spell

Many bugs in life can be solved by merely paying attention to them. The most important TAP to install is a meta-TAP, or Sapience Spell, that wakes you up

periodically into Kernel Mode and reminds you to pay attention.

Here's how to learn the Sapience Spell:

Finding the right trigger is of utmost importance. Treat this step with the gravity that a wizard puts into picking his wand or familiar.

The trigger should be concrete and constant in your life. Ideally, an item on your person at all times of sentimental value: a ring, a watch, a tattoo, a mole or birthmark on your hand, a specific gesture you make regularly. If not, it can be an attractive picture or bauble on your desk. Take your time to pick one that feels meaningful.

Once you've picked the trigger, it's time to pick the action. The action should be a mental move in the category of *pay attention*, but personalized: *breathe*, *reflect on my goals, be present in the moment, collect myself.*

Now, set yourself a Yoda Timer for 5 minutes and practice the Sapience Spell with the five steps above. Walk around as if doing your daily thing, notice your trigger, and rehearse the action. Do that ten times. Visualize yourself in different situations where a Sapience Spell would help. Let your mind wander a bit and then snap yourself back with the Sapience Spell.

For my own Sapience Spell, I picked a mole on the inside of my right thumb that I've had since childhood. After staring at it for some time and injecting feelings of attentiveness and intentionality, I find that notice where it is in physical space even without looking at it. I hope it proves a constant and comforting note in the future.

3. One Concrete TAP

If you have any kind of habit or routine, you're doing TAPs already. Today we will build one concrete micro-habit with TAPs.

Pick the easiest bug on your Bug List that might be solved by some kind of regular action. For example, I picked "forgetting things when I leave home."

Set a Yoda Timer for five minutes to design and install a TAP to solve that bug using the checklist in TAPs 101 above.

Reminders:

1. It's best to pick natural, concrete TAPs that you notice already. For example, I pay a great deal of attention to boundaries and thresholds. The trigger I picked is "stepping across the threshold of my apartment." Another example too good not to mention: I frequent a restaurant on campus called **T**he **A**xe & **P**alm. Every time I go I reflect on all the TAPs I'm currently installing.
2. Be realistic and pick baby steps for actions. Difficult habits should be built out of multiple TAPs. If your TAP is "After brushing my teeth, go running," make it "After brushing my teeth, take a walk" or "After brushing my teeth, go outside" or even "After brushing my teeth, look at the front door."
3. Keep rehearsing/practicing the TAP until five minutes are up. Yoda Timers quickly remind you *just how quickly you give up.*

Keep building one TAP a day over the course of Hammertime. If you're lost for ideas, try extending solidified TAPs into longer sequences of actions, one step at a time.

Soon you’ll have complete routines for specific situations. We’ll check in again on TAP progress on Day 13.

Daily Challenge

If you don’t mind, share your Sapience Spell.

Hammertime Day 4: Design

This is part 4 of 30 in the Hammertime Sequence. Click here for the intro.

A central theme of Hammertime is that rationalists interact with reality first. We try for at least five minutes. We build habits to solve the bugs. We stick our necks out and ask reality for rapid feedback. Only after doing our due diligence and getting beaten back by reality should we turn to introspection.

That's why the first five Hammertime techniques are for getting out and solving problems immediately. Only after interacting with reality and actually trying do we get to turn inwards to meditate, to question our motives, to get in touch with our feelings, and to make long-term plans.

Design is the most subtle approach for directly solving problems. It's about permanently distorting the physical reality around you to propel you towards – instead of away from – completing your goals.

Day 4: Design

Design (sometimes taught as Systemization at CFAR) is rationalist fengshui. Its main goals are:

1. To inject intentions into the physical space you live and work in.
2. To free attention from unnecessary and repetitive distractions.

Design principles work across domains: in the establishment of routines, in the shaping of social environments, and in the organization of screen space. For the first cycle of Hammertime we will focus on Designing physical space to make concrete and immediate improvements.

Here are the three core principles of Design, according to yours truly.

1. Intentionality

The very first tear I shed at CFAR was in response to Valentine's speech in Design class about the subtle machinations of Moloch infecting the very space around us:

The counter-top next to the front door that attracts mountains of junk like a gravitational well.

The dressers that hide away our running clothes and with them our good intentions.

The disarray that sends us stumbling back and forth to find our glasses, wallet, keys, and phone in the morning.

Moloch's servants appear wherever attention is lacking.

The first principle of Design is Intentionality: *things are where you intend them to be*. Look around your room or desk. Everything should have a purpose. The purpose can be functional, but it also can be sentimental or aesthetic. You can intentionally

organize things in the order you use them. You can intentionally arrange things in a pretty way. You can also intentionally leave a mess because you aspire to discover the next Penicillin. Regardless of how things are arranged, they should be arranged so *because that's what you intended.*

2. Amortization

The second principle of Design is Amortization: *pay up-front costs now to save attention in the long run.* Amortization is particularly concerned with the intentional placement of commonly used objects. Here are some examples to illuminate this principle:

I spent a day noticing all the wasted attention in my life. The first thing I noticed is that I search for my glasses all the time. Every time I wake up, come back from running, or get out of the shower, it'd take a minute to find them. The problem with having abysmal vision is that I can't see my glasses until they're right in front of my nose. The problem is exacerbated by the fact that my wife's glasses have thicker, more visible frames.

To solve the glasses problem, I picked a central location, a Schelling place if you will, to always leave my glasses, and placed my glasses case there permanently. Then, I rehearsed the TAP of taking my glasses off and putting them in the case. Four days later, it's become routine.

A number of other changes follow this principle: placing my keys and wallet in a box near the front door. Putting my running shorts on an easily reachable wall hook (you need more of these). Moving the oatmeal right next to the stove. Placing the vacuum cleaner next to its outlet.

3. Reflexive Towel Theory

Derived from Hitchhiker's Guide to the Galaxy, Towel Theory is an extension of the Fundamental Attribution Error which says that people decide *the kind of person you are* from superficial signals:

> A towel is just about the most massively useful thing any interstellar Hitchhiker can carry. Partly it has great practical value. [...]
>
> More importantly, a towel has immense psychological value. For some reason, if a strag (strag: nonhitchhiker) discovers that a hitchhiker has his towel with him, he will automatically assume that he is also in possession of a toothbrush, washcloth, soap, tin of biscuits, flask, compass, map, ball of string, gnat spray, wet-weather gear, space suit etc., etc. Furthermore, the strag will then happily lend the hitchhiker any of these or a dozen other items that the hitchhiker might accidentally have "lost." What the strag will think is that any man who can hitch the length and breadth of the Galaxy, rough it, slum it, struggle against terrible odds, win through and still knows where his towel is, is clearly a man to be reckoned with.
>
> Hence a phrase which has passed into hitch hiking slang, as in "Hey, you sass that hoopy Ford Prefect? There's a frood who really knows where his towel is."

The third principle of Design is Reflexive Towel Theory: *we all apply towel theory to ourselves*. Look at the space around you. It is telling you something about who you are. The blank walls call you a minimalist. The two-story rack of shoes reminds you how superficial you are. The unmade bed and unkempt piles of laundry, mail, and dirty dishes say, *you're not the kind of person who deserves to be cared for.*

Pay attention to what the space says about you – and not to other people, to yourself. Figure out if those are messages you want to be hearing. Maybe you want to hang up a print of Kandinsky's Composition 8 to replace that old Death Note poster. Maybe you'd like to be the kind of person who makes their bed. Whatever subliminal message your surroundings are sending via Reflexive Towel Theory, make sure it's the message you intend to hear.

Design Time

Today's exercise will take 10 minutes. Specify the physical space you'd like to redesign: anywhere from a single room to an entire house. Get pen and paper.

Step 1. Set a Yoda Timer. Walk around to explore the space and jot down all the things you'd like to change. Is there visible clutter you'd like to store away? Is there unsightly blank space? Is the furniture arranged in a productive way? Is there a better way to arrange commonly used objects to save time? What objects – appliances, decorations, furniture – are missing?

Step 2. Set another Yoda Timer. Hit as many things on your list as you can in that time. Move things and furniture around to their optimal permanent locations. Order all the organizational knick-knacks you need from Amazon. Pack the clutter away in empty suitcases, cupboards, or under the bed.

Daily Challenge

Fix as many bugs on your Bug List as you can by only rearranging physical objects. What's the hardest bug you fixed this way?

Hammertime Day 5: Comfort Zone Expansion

This is part 5 of 30 in the Hammertime Sequence. Click here for the intro.

It would be hypocritical of me to write a post of my usual form to teach Comfort Zone Expansion. Instead, I'll explain why the Disney song How Far I'll Go is a triumphant call to exploration, and leave a short CoZE exercise that you should modify with the principles of Moana in mind.

Background

Comfort Zone Expansion (ironically named CoZE) is CFAR's version of exposure therapy, designed to get people to try new things cautiously. When I first heard of CoZE, what came to mind was something like *run naked into a crowded Starbucks and ask strangers to finger-paint my buttcheeks*. Although there might be some value to such an exercise, CoZE is decidedly not that. The first step of CoZE is simply trying things you've never bothered to try, even though you have no resistance to them.

Let me call attention to some metaphors for talking about Comfort Zones.

Order and Chaos

One way to visualize your comfort zone as the dividing line between Order and Chaos.

Order is the known. Order is your social circle, the interior of your home, the streets you drive regularly. Order is the programming languages you're familiar with, the sports you play, the languages you speak. Order is the rules you follow. Order is your comfort zone.

Chaos is the unknown – or worse the unknown unknown. Chaos is staring momentarily into a stranger's eyes. Chaos is the antsy feeling you get turning just one street away from your usual route. Chaos is the feeling that *the world has shifted beneath your feet* when you break your code, when you find out you've been lied to, when you notice you're deep into a mistake. Chaos is the amorphous shadow that expands gas-like to fill every space you don't pay attention to.

Yang and Yin are Order and Chaos, and the Yin-Yang is the Daoist reminder that the proper Way through life is to navigate the twisting line between Order and Chaos.

For a more CS-friendly metaphor, consider staying within Order as Exploiting well-understood strategies and going into Chaos as Exploring new strategies. Moloch is the civilizational disaster that occurs everyone decides to Exploit by sticking within their comfort zones. Except for very young children, people categorically Explore too little and stagnate in local optima.

The Structure of Pop Songs

Jordan Peterson had an illuminating dialogue with composer Samuel Andreyev about a year ago (transcription my own):

> Andreyev: The pop song is an incredibly difficult medium to work within because – first of all it's completely unforgiving, you're working in an extremely compressed format, it's very rare for the pop song to be longer than three minutes. You don't have much room to maneuver. And you certainly don't have any room to maneuver structurally, I mean you pretty much have to stick to the verse-chorus-verse-chorus thing in the immense majority of pop songs, there's been very little variation of that since Rock really, since the Fifties.
>
> Peterson: Where did that come from? I know the three minute length was a commercial imposition if I remember correctly. But that structure verse-chorus-verse-chorus out of what did that originate?
>
> Andreyev: Well that's an extremely old form. Well you certainly have Baroque forms that have an extremely similar form. You alternate one fixed element that keeps returning the same way essentially and a secondary element that gives you a certain degree of relief and contrast with the preceding element.
>
> Peterson: So that's a Chaos-Order interplay of sorts, that's the way I would interpret it.

The verse-chorus-verse-chorus format of pop songs, then, is an alternation of Explore-Exploit as the song wends its way between Order and Chaos. The chorus is the primary, fixed element of Order that returns to tie the listener back to a central theme or narrative. The interspersed verses are exploratory elements that make brief forays into Chaos to provide relief from the monotony of the chorus.

This explains why other genres of music, less vernacular and more artistic, are less palatable to the public imagination. The avant-garde artist is the dedicated Explorer, constantly far into the lands of Chaos. Without a comforting refrain to return to, the music becomes all Chaos to the uninitiated and difficult to digest.

Moana

If you haven't already, take a listen or ten to How Far I'll Go. I personally prefer Alessia Cara's rendition.

> *I've been staring at the edge of the water*
>
> *'Long as I can remember, never really knowing why*
>
> *I wish I could be the perfect daughter*
>
> *But I come back to the water, no matter how hard I try*
>
> *Every turn I take, every trail I track*
>
> *Every path I make, every road leads back*
>
> *To the place I know, where I can not go, where I long to be*
>
> *See the line where the sky meets the sea? It calls me*

And no one knows, how far it goes

If the wind in my sail on the sea stays behind me

One day I'll know, if I go there's just no telling how far I'll go

I know everybody on this island, seems so happy on this island

Everything is by design

I know everybody on this island has a role on this island

So maybe I can roll with mine

I can lead with pride, I can make us strong

I'll be satisfied if I play along

But the voice inside sings a different song

What is wrong with me?

See the light as it shines on the sea? It's blinding

But no one knows, how deep it goes

And it seems like it's calling out to me, so come find me

And let me know, what's beyond that line, will I cross that line?

The line where the sky meets the sea? It calls me

And no one knows, how far it goes

If the wind in my sail on the sea stays behind me

One day I'll know, how far I'll go

Home to deep-dwellers and Lovecraftian horrors, the ocean has always been symbolic of Chaos. Moana teaches us three important methods of venturing into Chaos, all of which should be combined for maximum effect.

1. The Edge of the Water

The edge of the water is the line between Order and Chaos, constantly shifting with the lapping waves and the tidal cycle. The simplest method of CoZE is to stare the edge of the water and dip your toes in. That's what today's exercise is about. Everyone has a boundary they're drawn to inevitably, never really knowing why. The trick is to notice that boundary.

Every turn, trail, path, and road leads back to the edge of the water. Finding it is as simple as listening for the quiet yet shrill notes of resistance that stop you in your tracks in everyday life. The errand you put off for another hour. The acquaintance you almost wave hi to. The question you almost ask. The conversation topic clinging

desperately to the tip of your tongue as you try to launch it. The class or club you almost sign up for.

Life leads you back to the edge of the water no matter how hard you fight it. You've been staring at it for as long as you can remember. All you have to do is notice.

2. Where the Sky Meets the Sea

Staring at the very boundary between Order and Chaos may be useful for finding your resistances, but it's hardly a triumphant call to action. Moana reminds us to look up, every so often, at the line where the sky meets the sea. The sky is the Kingdom of Heaven, and it can only be reached by sailing farther out of your comfort zone than anyone has ever been.

There's an array of visual metaphors for successful, interesting people. They seem to be filled to the brim with the light of life. The light shines through them. They walk in the light of God. The second method of CoZE is to raise your eyes momentarily to meet that blinding light on the sea that marks your transcendent dream.

See the people you admire who shine and sparkle with the light. Construct the ideal human being in your mind's eye. Then, you will know what you're missing lies beyond the edge of your comfort zone. Let that transcendent dream be the wind at your back on the open sea.

3. Everybody on this Island

Why is Moana the only person on the island who yearns for the ocean? Is it because others are too afraid of their resistances, or cannot see the light on the horizon?

Actually, the reason Moana wants to leave is that she's already at the top of the dominance hierarchy on her island. She's the daughter of the chief, and she's destined to lead and been trained for it since childhood. Listen to her voice when she sings: "I can lead with pride, I can make us strong, I'll be satisfied if I play along." There's not a single note of worry or insecurity. Unlike everyone else on the island, the only way for Moana to grow is to leap out into Chaos.

This leads to the counter-intuitive third method of CoZE: expand the boundary of your comfort zone by securing the center.

Fortify and build trust within your relationships. Study and perfect your craft. Use Design principles to create a sanctuary to return to. Climb to the top of your current hierarchy. Once the center is secure, there will be nothing left for you here. Your natural inner voice will take you back on the open sea.

Day 5: CoZE

For the first cycle of CoZE, we will spend about half an hour trying new things.

WARNING: Don't pick anything you feel significant resistance to. The goal is simply to become the kind of person who automatically tries new things if they're nonthreatening.

Step 1. Set a Yoda Timer for five minutes. Brainstorm as many things you haven't done as you can. They can be as simple as: listen to songs in different languages, walk down a street you haven't been down before, try to do a handstand against the wall, shout as loud as you can, run a mile, have a conversation without smiling, write a haiku.

Step 2. Set a Yoda Timer for TWENTY minutes. Hit as many of the things on your list as you can.

Daily Challenge

Share a story about finding something shiny by exploring past your comfort zone.

Hammertime Day 6: Mantras

This is part 6 of 30 in the Hammertime Sequence. Click here for the intro.

I'd like to demarcate the line between the two natural halves of Hammertime (fast and interactive vs. slow and introspective) with an experimental post, more reflective than actionable.

Motivation

The seed of this post was planted in my mind by a conversation with Zvi. In said conversation, he invited me to read the rulebook of Mage: The Ascension and take it as literally as possible. The particular magic mechanic that struck me from Mage was the Paradox phenomenon, which (roughly speaking) causes magic to backfire in the presence of Muggles.

> When it is performed ineptly, or is vulgar, and especially if it is vulgar and witnessed by sleepers, magic can cause Paradox, a phenomenon in which reality tries to resolve contradictions between the consensus and the Mage's efforts. Paradox is difficult to predict and almost always bad for the mage. The most common consequences of paradox include physical damage directly to the Mage's body, and paradox flaws, magic-like effects which can for example turn the mage's hair green, make him mute, make him incapable of leaving a certain location, and so on. In more extreme cases paradox can cause Quiet (madness that may leak into reality), Paradox Spirits (nebulous, often powerful beings which purposefully set about resolving the contradiction, usually by directly punishing the mage), or even the removal of the Mage to a paradox realm, a pocket dimension from which it may be difficult to escape.

The upshot is not dissimilar from the rather commonplace observation that extraordinary people seem to distort the reality around them and also have a difficult time imparting their reality-distortion field to others.

My foray into the fantastical world of Mage led me to consider taking other mechanics of magic more seriously. Among the infinite variety of ways the human mind might break the laws of physics, only a very few magic mechanics have lasted in the public imagination. Again and again, fantasy writers return to the incantation: words that effect transformation by their mere utterance. What is so psychologically fascinating about incantation?

And if a single utterance can effect magic, what might the repetition of words of power accomplish, over the course of many years?

Day 6: Mantras

Epistemic status: true story.

I was not a particularly well-socialized child growing up, but even in sixth grade I knew there was something *wrong* with her. She was a bit standoffish, her hair a bit too bushy and disheveled, and she spoke with the cadence of a lost soul. If she had a

name, it must have been something like Elphaba. I couldn't place it at the time, what exactly was *wrong* with the girl. Only now, more than a decade later, can I give a name to her intensity: that uncommon ability – inimical to the instincts of all sixth graders who desire to fit in – to take ideas seriously.

I only ever had one conversation with the girl. I don't recall what class it was in – some discussion group, perhaps, for a play of Shakespeare's far above our reading level. While the teacher popped out to grab dry-erase markers, some eight of us sat around the round discussion table fidgeting as sixth graders are wont to do.

Then, somehow, the girl to my left transfixed me with her gaze and spoke:

Girl: *Memento mori, memento vivere.*

Me: Sorry?

Girl: It means, Remember that you are going to die. Remember to live.

[long pause]

Girl: *Memento mori, memento vivere*.

Surely, such a conversation must have been bracketed by benign chatter. Perhaps I triggered it with a bout of adolescent nihilism. Perhaps we led up to it by a meditating on "To be or not to be?" or "Alas, poor Yorick!" Then again, knowing that girl, perhaps not.

I never saw her again. As far as I know, she completely vanished after the sixth grade.

Memento mori, *memento vivere*.

I cannot say how many years those words haunted me. I can say, however, that in the dark of countless middle school nights I was tormented by the shadow of mortality. That in the light of day *memento vivere* stirred in my heart a frantic energy to rise to the occasion and battle the injustice of being itself. That I repeated these words under my breath as I pondered questions of philosophy such as "Does teleportation kill the original copy of you?"

That half a decade later, when the girl's voice had subsided into distant memory, I decided for some inarticulate reason that *Memento* was my favorite film before its opening credits finished rolling.

How many years did *memento mori* haunt me? You might say that the rest of my life, from the point of that conversation, has been a quest to recover words of such power from thousands of novels, manuscripts, songs and videos. Mantras to remind me of the direction of my transcendent dream. They speak now.

Everything can be made radically elementary.

That which can be destroyed by the truth, should be.

People can stand what is true, for they are already enduring it.

The purpose of mathematics is to advance human understanding.

People become who they are meant to be by doing what is right.

Modern people cannot find God because they will not look low enough.

The line between Good and Evil runs through the heart of every human being.

But I, being poor, have only my dreams. So I spread my dreams beneath your feet. Tread softly on them.

There's a naive rationalist in me calling me a sucker for falling hook, line, and sinker for Deep Wisdom. To him I have only this to say: whereas by repeating these mantras I am filled with a renewed energy and direction for life that lasts for many years, he would have come away with only a vague cynical smugness. So who's winning now?

I have the sense that the mantras I repeat under my breath are imbued with my deepest values and serve as a solution to the Control Problem in myself: to cheaply propagate those values to future copies of me across the span of many years.

Daily Challenge

Share a favorite mantra and what it means to you.

Hammertime Day 7: Aversion Factoring

This is part 7 of 30 in the Hammertime Sequence. Click here for the intro.

As we move into the introspective segment of Hammertime, I want to frame our approach around the set of (unoriginal) ideas I laid out in The Solitaire Principle. The main idea was that a human being is best thought of as a medley of loosely-related, semi-independent agents across time, and also as governed by a panel of relatively antagonistic sub-personalities à la Inside Out.

An enormous amount of progress can therefore be made simply by articulating the viewpoints of one's sub-personalities so as to build empathy and trust between them. This is the aim of the remainder of the first cycle.

Day 7: Aversion Factoring

Goal factoring is a CFAR technique with a lot of parts. The most resonant sub-skill for me was Aversion Factoring, so we'll start there. I highly recommend Critch's TedX talk on the subject, where I first learned this way of thinking.

Setup

Pick from your Bug List a habit you want to start but haven't, or that you've been forcing yourself to do but remains a drag. What's happening?

For concreteness, let's say the habit is "blog every day."

At some level, you want to blog. You have good ideas. Writing helps you think clearly. You'd reap the benefits of being publicly wrong. If you blogged, other human beings might benefit. But if you really *wanted* to blog then why does it cost so much willpower every time? Why aren't you leaping into it every day the way you leap into deep fried ice cream?

Aversion factoring is about noticing and removing the subconscious roadblocks keeping System 1 from wanting the same things System 2 wants.

1. Articulate Aversions

The first step to Aversion Factoring is to articulate the aversions that are holding you back. Begin by listing all the reasons you don't like about doing the thing. Two things to keep in mind:

Be honest.

"I'm afraid my ideas aren't original, my writing hasn't improved since fifth grade, and I'm terrified of people on the internet."

Being honest is difficult. However, there's a second category of insidious aversions: trivial, repetitive annoyances that leave a bad taste surrounding the whole experience. See Beware Trivial Inconveniences. Finding such aversions requires attention to detail:

"I hate blogging because of the awful LaTeX support, because every time I want to include a picture I get anxious about copyright issues, and because I recently discovered a popular blogger friend has the exact same WordPress template so if I change mine I *lose* and if I keep it the same I feel like a copycat so I'd rather just block the thoughts out aggghhhh."

The primary focus of today's exercise is to find and debug the trivial inconveniences in our lives.

2. Decide Whether to Endorse the Aversion

For any given aversion, there are two ways to proceed. Endorse an aversion if it points to a real underlying problem that needs to be solved. In my blogging example, I might decide that I care about writing quality and targeted writing practice is long overdue.

If you don't endorse the aversion, then it's unnecessary and should be removed. A common class of such "bad" aversions is bucket errors about identity. When deciding to remove aversions, remember Chesterton's Fence! Figure out why you have the aversion before you try to remove it. Almost any aversion can be removed by gradual exposure, so be careful (see Boiling the Crab).

3. Solve or Reduce Aversions

Once you've figured out what the aversions are, it's time to solve them as much as possible, one by one. For endorsed aversions, the course of action is to modify or upgrade the habit itself to solve or sidestep the underlying problem. To solve my writing problem, I might decide to reread and act on Strunk and White or Nonfiction Writing Advice. (Huh. That's a good idea.)

Meanwhile, un-endorsed aversions should be targeted with exposure therapy or CoZE. To apply exposure therapy, build a path of incremental steps towards the aversion, each of which feels individually safe. Take steps one at a time as gently as necessary. I gently amped up my blogging frequency over about a year to an audience of zero, then one, before I got over my fear of y'all internet people.

CoZE is the upgrade to exposure therapy in which you build in ejector seats: pre-commit to multiple points along the route where you can reflect on whether or not you endorse the aversion.

Aversion Factor Three Bugs

For today's exercise, please pick THREE bugs from your Bug List related to habit-building. These can be habits you want to pick up, or habits you already have but want to upgrade.

For each bug, set a Yoda Timer for five minutes and Aversion Factor it:

1. Walk through the habit and list out as many aversions as you can, paying particular attention to trivial inconveniences.
2. Decide for each aversion whether or not to endorse it.
3. Solve as many as you can in the remaining time.

Daily Challenge

I have a friend who stays in bed for hours in the morning because it’s too cold to make the voyage across his bedroom for clothes. Share a trivial inconvenience in your life that might have (or has had) dramatic consequences.

Hammertime Day 8: Sunk Cost Faith

[Author's note: I will be moving future Hammertimes to my personal page to avoid cluttering the frontpage. This one is sufficiently short and probably controversial to leave here.]

This is part 8 of 30 in the Hammertime Sequence. Click here for the intro.

It pains me to begin a post about planning with an announcement about two slight changes of plans for Hammertime:

First, I will be travelling the week after next, so there will be a week-and-a-half intermission between the first and second cycles.

Second, when I sat down to write a post about Focusing, I found myself unable to add anything productive to this excellent post: Focusing, for Skeptics. Focusing is probably the second most powerful technique I learned from CFAR, so I will return to it in future cycles after more thought.

Instead, I want to write three posts on planning. These are the first steps to becoming the kind of person who can make thoughtful long-term plans and follow through with them.

Day 8: Sunk Cost Faith

One of my main motivations ever since writing The Solitaire Principle is to solve the Control Problem in humans: the problem of making and following through with long-term plans and habits despite new information and, even worse, value drift. I propose that what is commonly known and vilified as the Sunk Cost Fallacy actually exists for a good reason and is a useful first-order solution to the Control Problem.

The Uncanny Valley of Sunk Cost

Related: Sunk Costs Fallacy Fallacy.

Here is the uncanny valley one falls into when one naively cancels out Sunk Cost Fallacy:

1. You suck at making plans, but follow through with them anyway. You get a moderate amount done by making overconfident and poorly-thought-out plans, and just doing them despite contradictory information.
2. One day, you learn about the Sunk Cost Fallacy. You decide to be a Good Rationalist and categorically abandon ship on projects that no longer appeal to you. You still suck at making plans. All your plans fail and you get nothing done.
3. Over time, you learn that you're not the kind of person who can follow through with long-term projects. You jump from bright light to bright light, captured by the briefest caprice. You don't remember what it's like to be diachronic. Your time horizon shrinks and you stop bothering with plans at all.

There's an extremely insidious demon hiding at Step 2, related to adverse selection. Over the course of a multi-year (or multi-day, for that matter) plan, all sorts of noisy information can arise. Imagine your valuation of the project to be something like a Brownian motion bouncing around due to new information that slowly converges towards the "true value."

If *at any point* the current valuation of the project randomly walks under the Worth It line, you'll promptly give up the project.

Because of the noisiness of information, following the strategy of "give up whenever it falls below the Worth It line" will make you give up on many projects that actually turn out to be Worth It, just because a long random walk will always usually fall quite a bit below the mean at least once.

And this doesn't even take into account all the motivated reasoning and other reasons that shiny new projects putter out of momentum.

Sunk Cost Faith

I think the Uncanny Valley above is a serious and common failure mode in the rationalist community, and one that happened to me.

My diagnosis is that one should not fix one's Sunk Cost Fallacy without first learning to make strong, fault-tolerant plans, and that one cannot learn to make strong, fault-tolerant plans without the data that comes from following through on bad ones. Therefore, the first step towards becoming good at planning is restoring your Sunk Cost Fallacy and use it to follow through on bad plans. This move I dub Sunk Cost Faith – faith that your past self made good decisions. Faith, of course, because it's entirely unjustified.

If you find yourself in the position described in the previous step, pick up your Sunk Cost Fallacy again and turn it into Sunk Cost Faith. Build yourself into a diachronic person. Follow through on your plans even after they no longer appeal to you. Expand your time horizon to the scale of months and years so that you're the kind of person who can actually do things.

Once you can actually follow through with plans again, only then can you get better at planning. This will involve explicitly building into your plans defenses against the dark side of Sunk Cost Fallacy, defenses such as unambiguous ejector seats.

Faith in the Past

Today's exercise is directed at people who find themselves giving halfway on projects too frequently.

Pick a completely useless activity (be creative!) that takes about five minutes, and do it every day for a week with Yoda Timers.

Daily Challenge

Convince me that I'm wrong about Sunk Cost Fallacy and it's actually just bad.

Hammertime Day 9: Time Calibration

This is part 9 of 30 in the Hammertime Sequence. Click here for the intro.

I've been thinking about whether or not regular betting, prediction markets, and being well-calibrated is actually useful, and if so how to practice train calibration on a short feedback loop.

Being able to make accurate time-to-completion estimates, at least, is extremely powerful. This post describes my current strategy for staying calibrated about time estimates.

Day 9: Time Calibration

Of all the cognitive biases in the Sequences, Planning fallacy seems to be one of the most directly harmful and eminently fixable. The goal of today's exercise is to build a tool for routinely checking your calibration about how long things take.

Although Planning fallacy is the clear antagonist in this situation, I also want to gesture at a second class of failures I've been facing which involve systematically *overestimating* the difficulty of things.

Vortices of Dread

After spending a few days checking my calibration, I was surprised at the sheer number of things I routinely overestimate the difficulty of, mostly because of an ingrained fear of housework and the bureaucratic machine.

Several years ago, I watched my father spend nearly a full week on taxes, reading all the fine print, cross-referencing internet forums, and triple-checking every field. I did my own taxes for the first time last year, expecting it to only be more nightmarish – after all, my father's experience surely saved him mounds of time already, right? Instead, the whole process took me a single afternoon.

Two weeks ago, I set out to get a travel visa done with a travel agent after months of dread. I blocked off the entire afternoon in the (it seemed to me) likely event that I'd have to drive back and forth to pick up, print, and/or fix documentation. The visit ended up taking a total of ten minutes, not counting the two mile drive.

Last week, I set out on an odyssey to make a photo album for family members, dreading the many evenings I would pore over old files and spend arranging prints. The whole process took two and a half hours from start to finish with the timely aid of Yoda Timers.

What threw my award-winning calibration off so wildly? Two things were at play:

First, most of my System 1 data on how long things take comes from watching my anal-retentive parents. I instinctively feel that cooking a meal takes nearly an hour, that every field on every form needs to be checked twice by every individual involved,

that you should always arrive fifteen minutes early, and that the bureaucratic machine is constantly out to Get You. That gave me the opposite of Planning fallacy.

Second, the creeping dread around problems became a self-fulfilling prophecy. Even though I was relieved after finishing my taxes in one afternoon, my memory of the experience is still dominated by the weeks of slowly building anxiety leading up to the event. In contrast, I hardly remember the actual filling out of forms. I suspect System 1 was picking up these awful weeks as signal of doom.

Calibration Challenge

In the next day, keep an eye out for clear-cut work and train your calibration on how long things take. Make a System 1 prediction about how long each activity takes, set a Yoda Timer for that amount of time, and try to hit that time. If you expect something to take longer than an hour, break it up into clearly-demarcated chunks to calibrate individually.

(There is, of course, the confounding factor of the timer, but if you find yourself significantly more efficient with the timer goading you on … maybe that’s something to consider doing regularly.)

If you’re anything like me, you’ll be wildly surprised by how systematically wrong your models are, in at least one direction. If surprised, update!

Daily Challenge

Share your worst case of Planning fallacy.

Hammertime Day 10: Murphyjitsu

This is part 10 of 30 in the Hammertime Sequence. Click here for the intro.

> Like, so pessimistic that *reality* actually comes out better than you expected around as often and as much as it comes out worse. It's actually really hard to be so pessimistic that you stand a decent chance of *undershooting* real life.

Later in the day I will put up an open thread about the first cycle of Hammertime.

We finish up the first cycle with another post on planning. Murphyjitsu is CFAR's method for planning which asks us to try to be so pessimistic as to undershoot real life.

Day 10: Murphyjitsu

Murphy's Law states that anything that can go wrong will go wrong.

For our Mandarin-speaking readers, here's a useful mnemonic: Murphy transliterates as 墨菲 (*mo fei*), which is homophonous to 莫非, "what if?" That's why I think of Murphy's Law as the What If Law.

In the course of making plans, Murphyjitsu is the practice of strengthening plans by repeatedly envisioning and defending against failure modes until you would be *shocked* to see it fail. Here's the basic setup of Murphyjitsu:

1. Make a plan.
2. Imagine that you've passed the deadline and find out that the plan failed.
3. If you're *shocked* in this scenario, you're done.
4. Otherwise, simulate the most likely failure mode, defend against it, and repeat.

The first important sub-skill of Murphyjitsu is Inner Sim – the ability for System 1 to simulate failure modes.

Inner Sim

I have the suspicion that everyone is secretly a master at Inner Sim, the ability to instantly simulate failure. Imagine a friend declares to you their New Year's Resolution: to write a novel, to go on a keto diet, to write a month-long sequence on instrumental rationality.

Now, listen for that internal scoffing – your System 1 instantly proliferates the future with all manner of obstacles. That's Inner Sim at work.

If you're anything like me, Inner Sim is better at predicting other people's failure modes than your own. The mental move that helps apply Inner Sim introspectively is essentially Outside View: take your plan and imagine another person made them. What will go wrong?

Welp Mentality

Inner Sim does surprisingly little on its own.

I had a conversation with a rationalist friend (let's call him "Alex") that went something like this:

Alex: What's bothering you?

Me: I've been terribly unproductive. I'm procrastinating on fellowship essays … they're due in two weeks, and every time I think about math these essays pop up in my head.

Alex: Why?

Me: I essentially finished them, but I still have to edit it. Copy-editing is so tedious, and every run I make through my writing, it looks even more awkward than it did the previous time.

Alex: What do you predict will happen?

Me: Well … I'm going to put the essays off until two days before the deadline, edit for ten minutes when I start feeling the pressure, and submit them. Until then, I won't get any research done.

Alex: So…?

Me (shrugs): Sucks, right?

Alex breaks down in laughter.

I call this Welp Mentality. Welp Mentality is noticing that your plans are likely to fail catastrophically, or run overtime, or take 10x as much effort as you thought, and then shrugging noncommittally. Welp.

Welp Mentality is knowing and accepting as a fact of life that every build will release two months late. That you'll end up half-assing problem sets and essays starting midnight before the deadline. That your current exercise plan will probably peter out. I had an old motto for Welp Mentality: "Due tomorrow? Do tomorrow."

Murphyjitsu

Murphyjitsu is the astounding notion that if you can predict a failure mode, you can *do something* about it!

If your builds release two months late every time, you can move the release date, or cut features, or hire more engineers. If you know you're only going to spend six hours on a problem set the night before the due date, at very least you might as well just set a six-hour Yoda Timer for it, do it at a convenient time, and submit whatever you end up with.

In my fellowship essay case, I decided to spend ten minutes editing and submit the thing immediately. The relief of getting two weeks of my life back was palpable.

Pick a plan you have for the near future. Murphyjitsu it. Pull out all the stops: Arrange social pressure to keep you on track. Double the amount of time you spend. Set calendar and phone reminders. Murphyjitsu only stops when you would be *shocked* if the plan fails.

Daily Challenge

Murphyjitsu a central life goal. Are there glaring failure modes you haven’t defended against?

Hammertime Intermission and Open Thread

This post marks the end of the first cycle of Hammertime. Click here for intro.

Hammertime will return on Monday 2/19.

I want to close off the first cycle with some thoughts, and designate a place for discussion about the future of this sequence.

Discussion Topics

1. Sequences: Yea or Nay?

I've always felt that sequences are a valuable way to organize deeper thoughts and drive home a few central messages from several perspectives. However, the current format and culture on LW seem to radically favor short, independent chunks. (There is also the obvious problem that Sequence construction is not working.)

I've been posting daily for a while now but when I shifted from individual posts to a sequence, average Karma immediately dropped by a factor of about 2. It's possible that people don't bother upvoting the same sequence, or that my writing quality dropped, but if this is real signal that many more people would read a sequence if they are marketed as individual thoughts (and WordPress stats suggest this as well), that might be reason for me to stop writing sequences in the future, or at least collect sequences together only after they're complete.

Possible actionable for meta: make posts in a Sequence share karma and/or a single slot on frontpage.

2. Repeat or Explore?

My original intention was to review 10 topics over three cycles, building up in the difficulty of problems solved. I think I will definitely return to and expand on several of the techniques we've seen already, but also add more topics. If people have favorite techniques (and hopefully references) they'd like to see in Hammertime, post them here.

3. Monotonicity of Progress

A big goal of mine is to solve the "Rationalist Uncanny Valley," where beginning rationalists get worse at life before they get better. I can't believe that this has to be the case; it seems to be symptomatic of a larger failure to develop the proper curriculum. I would like progress on rationality to be monotone – is there a good reason this should be difficult? It'd be great if we could compile a central list of "uncanny valley" failure modes.

Bug Hunt 2

This is part 11 of 30 of Hammertime. Click here for the intro.

CFAR has an underlying mantra “adjust your seat”: systematically modify every technique and class to fit your personal situation. It’s common sense nowadays that different things work for different people, but the extent to which is true still constantly surprises me. (Kierkegaard had a fun take on adjusting your seat which he called the Method of Rotation.)

If you wish to partake in Hammertime, feel free to adjust your seat as much as necessary. Draw out the practice of instrumental rationality over a longer period of time, pick and choose the methods that appeal to you, and scale them to your time constraints.

Hammertime: The Second Cycle

Hammertime is about cultivating a tiny number of powerful techniques for solving a huge variety of problems. In the second cycle, we will revisit and upgrade the tools we introduced in the first, and apply them to tougher problems:

1. Bug Hunt
2. Yoda Timers
3. TAPs and Reinforcement Learning
4. Design
5. CoZE
6. Focusing
7. Cruxes
8. Goal Factoring
9. Internal Double Crux
10. Self-Trust

The new ideas we will be introducing in the second half are devoted to developing higher levels of introspection and self-honesty, to figure out your true motivations and aversions, and what to do about them.

Before each post in the second cycle, take a moment to review its predecessor.

Day 11: Bug Hunt 2

Previously: Day 1.

Noticing your bugs continues to be our single most powerful technique. Training noticing involves lateral thinking, attention to detail, and self-honesty. Today, we focus on three high-level ways in which human beings systematically err.

Setup

First, review your Bug List from Day 1 and update it.

For each of the next three mini-essays: read it over, then set a Yoda Timer for five minutes and brainstorm as many bugs as you can during that time.

1. Identity

Paul Graham wrote Keep Your Identity Small. Being attached to your identity can often constrain your growth.

Rather than making an impartial decision on what kind of person to be, people often extrapolate their identity (and morality) from their previous actions. A friend of mine calls this coprolite: fossilized and over-fitted beliefs that originate from early childhood. Are you neat or messy, stingy or generous, introvert or extrovert, conscientious or agreeable, idealistic or cynical, engineer or artist, vim or emacs? Do you look down on people for being the other way? Take moment to notice all the traits you're attached to, think about why you're attached to them, and consider the benefits of their opposites.

Personalities are many-faceted, and you may not even understand your true motives, fears, or skills. Do your stated preferences agree with your revealed preferences? Do your aliefs differ from your beliefs? Do people systematically judge you to be different from your self-image? Do you often surprise yourself in terms of what you enjoy, excel at, or are anxious about?

It's useful to think of personality growth as expansion rather than change. An introvert grows by learning how to navigate social scenes. An extrovert grows by reclaiming her capacity to be alone. Instead of asking what you would change about yourself, ask what you would add to the toolbox.

2. Pica

Pica is an eating disorder in which people crave food that don't fulfill the need behind that craving; the folklore example is gnawing on ice to satisfy a mineral deficiency. Experiential pica is any craving which doesn't fulfill the need behind it.

My top three addictions in high school were all experiential pica.

The first addiction was romantic novels and shows of a tragic nature, which served as vulnerability and sacrifice porn. I had intricate daydreams in multiple languages of love and loss.

The second addiction was RPG games, which served as improvement porn. In Diablo III, the Gem of Ease that boosts your leveling speed on all future characters to go from 1 to 70 in about an hour; I'd start a new character every couple months to get watch the level up messages roll in. MOBAs are perhaps the worst offender in this regard, taking your character from level 1 to fully equipped level 18 every single game.

The third addiction was just …

I know these are pica because the first and third cravings largely subsided when I entered a committed relationship, and the second when I started seriously working on self-improvement.

Lent is a good time to look for your pica. Are there any habits, cravings, or addictions you don't understand and/or try hard to cut? If they're pica, you're applying effort at the wrong angle. Figure out the unmet need, meet it, and the pica will automatically subside.

3. Ambition

I've been jogging casually for about fifteen years. Until last year, it's been uniformly awful. You'd think you'd get used to running four miles after doing it twice a week for a decade. You'd be wrong.

Then, I decided to aim at something.

I thought: *I'm going to train for a seven minute mile*.

My heart replied: *Oh, ok, that's kind of invigorating.*

Then I thought: *I'll train for a six minute mile.*

My heart: *Woo baby, let's do this!*

Then I thought: *A five minute mile.*

My heart: *HAHAHAHAHAHAHAHAHA…*

I ran for over a decade with next to no improvement. Last month I ran a seven minute mile after two months of training for an impossible goal. These days I look forward to running.

I've been blogging casually for about five years. Until last year, it's been a drag. You'd think you'd get better at writing by putting up two posts a month for a year or two. You'd be wrong.

Then, I decided to aim at something.

Me: *I'll try blogging once a week.*

My heart: *Oh, ok, that's nice.*

Me: *I'm going to blog every other day.*

My heart: *Now we're getting somewhere.*

Me: *I'm going to blog every day for a year, and write better than Eliezer Yudkowsky by the end of it.*

My heart: *HAHAHAHAHAHAHAHAHA…*

There's a level of ambition that pushes you to operate at maximal efficiency, that twists your heart with adrenaline just to think about. In every pursuit, aim at a target so high it feels immodest to whisper in an empty room.

List your goals now. Keep doubling them in difficulty until your heart bends over in hysterical laughter at the very thought.

Daily Challenge

State your greatest ambition: the one that feels most subjectively immodest.

Yoda Timers 2

This is part 12 of 30 of Hammertime. Click here for the intro.

> Anyone who can muster their willpower for thirty seconds, can make a *desperate* effort to lift more weight than they usually could. But what if the weight that needs lifting is a truck? Then desperate efforts won't suffice; you'll have to do something *out of the ordinary* to succeed. You may have to do something that you weren't taught to do in school. Something that others aren't expecting you to do, and might not understand. You may have to go outside your comfortable routine, take on difficulties you don't have an existing mental program for handling, and bypass the System.

~ *Make an Extraordinary Effort*

I don't know if I've ever made an extraordinary effort (that's probably evidence I haven't), but I've certainly made desperate efforts. The philosophy of Yoda Timers is that it might be enough to make desperate efforts all the time: to do the known thing as well and quickly as can be done. Past that is the realm of rare genius.

CFAR calls Yoda Timers "Resolve Cycles," a sub-skill of Resolve – the ability to make a desperate effort. Least glamorous of all rationality techniques, Resolve deserves its own book. How much could you accomplish just by *more brute force* all the time?

Day 12: Yoda Timers

Previously: Day 2.

Resolve is the main skill being trained by Yoda Timers, but there are a number of other useful reasons to build timers and deadlines into your life. Today I'll share three ideas to make the most out of Yoda Timers.

Yoda Deadlines

Sometimes, you surprise yourself with what can be done in five minutes. But sometimes, there are things that can't be done in five minutes. In this case, the generalization of Yoda Timers is to set absurdly short deadlines for these tasks.

How long does it take to write a novel? NanoWriMo is a Yoda Deadline for one month.

How long does it take to solve long-standing research problems? The IMO says: sometimes, only four and a half hours.

How long does it take to turn your life around? How many people waste away for years or decades before accelerating back through life in the span of weeks, tipped by a single conversation or book or trip?

The short answer to all these questions is: you have no idea how fast you can be without practicing for speed.

The Harvard-MIT Math Tournament (HMMT) has an easier version, the Harvard-MIT November Tournament (HMNT), which is run for local and less experienced (middle and early high school) students. HMNT is composed of several individual and team-based rounds, the most exciting of which is the Guts Round. Teams of 4-6 students work together on problems that come in sets of three to solve a total of 36 questions in 80 minutes.

A handful of older students, myself included, helped out at the HMNT of 2011. The coach of the IMO team challenged us to participate in the Guts Round, except instead of working in teams of 6 we would work alone, and without scratch paper.

And so it came to pass that, behind an auditorium full of teenagers loudly whispering ideas and trading scratch paper, the five of us sat silently in a row, staring problems down and writing down answers.

Tallying our scores at the end, each of us individually beat all of the actual teams by a wide margin.

From that day on, I did HMMT practice problems with half the time and only mental math. I won the thing twice.

The Race Against Decay

> "My dear, here we must run as fast as we can, just to stay in place. And if you wish to go anywhere you must run twice as fast as that."

~ Alice in Wonderland

There's a common failure mode with writing projects: if you work too slowly, ideas become stale before you're even close to finishing.

How many unfinished thoughts get the backspace because they don't stand up to reflective endorsement?

A half-finished blog post rusts overnight.

After a week, the first chapter of your novel reads like a child's writing.

That proof you jotted down months ago? You haven't a clue how to fill in the details.

I give examples in writing because I'm preoccupied with writing, but staleness and motivation decay apply to all creative pursuits, especially for episodic people. One solution is to try to solve the control problem, build trust with your future self, and otherwise learn to plan for the long term. That we covered on Days 8, 9, and 10. But another solution is simply to *do things faster*.

Murphyjitsu should have no trouble detecting these failure modes. There are ideas that you know you won't follow through on if you don't finish them immediately. If you put something off for months, even when you end up doing them, they'll take twice as much effort.

Set Yoda Timers and Deadlines. Motivations and values drift – make the most of those you have today.

Take it Slow

Usually, five minutes is an absurdly short time to try something. But sometimes five minutes is an eternity. The secondary use of Yoda Timers is to draw your focused attention to tasks that you normally spend seconds on.

How much time do you spend planning your day? Set a Yoda Timer and move things around on your schedule to maximize efficiency.

How much time do you spend expressing gratitude? Set a Yoda Timer and searching for the perfect gift, or writing a thoughtful note, for a loved one.

Are there muscles you never exercise? Set a Yoda Timer and train that one muscle group (see Sore in Six Minutes to learn how). Notice what flexing and relaxing it feels like. Explore the full range of motion. Accept the lovely burn of lactic acid.

Do you dive into things without enough planning? Set a Yoda Timer to slow down and Murphyjitsu.

Meta-Yoda

Today's exercise: set a Yoda Timer for five minutes and build a plan to incorporate timers and deadlines into your life.

Daily Challenge

Set a Yoda Timer and share the most important idea you haven't had time to express. Five minutes is all you get.

TAPs 2

This is part 13 of 30 of Hammertime. Click here for the intro.

> “Omit needless words!” cries the author on page 23, and into that imperative Will Strunk really put his heart and soul. In the days when I was sitting in his class, he omitted so many needless words, and omitted them so forcibly and with such eagerness and obvious relish, that he often seemed in the position of having shortchanged himself — a man left with nothing more to say yet with time to fill, a radio prophet who had out-distanced the clock. Will Strunk got out of this predicament by a simple trick: he uttered every sentence three times. When he delivered his oration on brevity to the class, he leaned forward over his desk, grasped his coat lapels in his hands, and, in a husky, conspiratorial voice, said, “Rule Seventeen. Omit needless words! Omit needless words! Omit needless words!”

~ *The Elements of Style*

There is nothing more essential to the practice of Hammertime than repetition, and no rationality technique that requires more repetitive practice than TAPs. Although we pick only three days to focus on them, it’s best to draw out the repetitive drilling of TAPs over a lifetime.

Day 13: TAPs

Previously: Day 3.

Triggers that Notice Themselves

The real skill with trigger-action planning is picking the right trigger. The best triggers are not only easy to notice, but hard to miss. It should not require effort and conscious attention to notice the trigger – the only conscious action occurs after the trigger calls the action to mind.

Three ways to find great triggers:

1. Sentimental value: there’s a process by which we naturally become attached to the items that accompany us through thick and thin. I am attached, for example, to the freckle on my right thumb, to a long-sleeve shirt gifted me by a childhood friend, to my Logitech gaming mouse – relic of a past life. Pay attention to these objects. Notice how they gain three-dimensionality. Inject them with meaning. For example, there’s a well of metaphysical space under my thumb freckle where I can store a preternatural calm for a minute of need.
2. Novelty: surprise is the easiest way to notice. Last month I made a number of purchases for Design, and their presence registered as unusual for weeks. Take advantage of new possessions to build micro-habits. My new welcome mat is a reminder to check my keys and phone before leaving the apartment. My new bean-bag chair tells me to notice and relax any muscle tension. My new reading lamp wants me to read every night before bed.

3. Felt sense (sneak peak for Gendlin's Focusing): a *felt sense* is a bodily sensory experience attached to an emotion or idea. Many powerful cognitive habits amount to building smart TAPs for specific felt senses. Most of the felt senses I notice center in my chest and spine. A twisting in my heart that tastes like Sour Patch Kids signals romantic feelings. A buzzing of energy that travels up my spine signals excitement (this one's probably just adrenaline). A physical pressing on the whole chest signals anxiety. Build plans for responding to each felt sense. Caution: suppressing is rarely the correct answer.

Sapience Spell Overloading

A general-purpose Sapience Spell has a large number of uses, and it's best to overload one trigger with them all. The Sapience Spell should trigger throughout the day: it will be clear from context which usage is most applicable.

Here's three new ways I've been overloading the Sapience Spell:

1. Refresh: A conversation is going nowhere. You're completely lost in an hour-long seminar on infinity-one categories. You lost sleep to AlphaGo nightmares and your whole day is fucked. You're half-way through a week-long conference on the Python compiler and entirely tapped out. What demonic presence compelled you to sign up for that? At every scale, the Sapience Spell can be the refresh button you need to clear sunk cost, memory leaks, and bad vibes. A "step back and relax" button for heated political conversation. A Ctrl-Alt-Delete for a clever but content-free blog post. A System Restart button for a project worth nothing but sunk cost. A FACTORY RESET (WARNING: ARE YOU SURE?) for triggering that mid-life crisis you desperately need.
2. Reality Check: I spent a week-long trip practicing lucid dreaming (to no avail yet). One of the main tools there is doing reality checks ("am I awake?") regularly, and that's now built into my Sapience Spell: look down at my hand and count my fingers. A reality check is a moment to notice your body exists and check for a bare minimum of sanity.
3. Reinforcement Learning: Check out Tune Your Cognitive Strategies. Use your Sapience Spell to regularly pat yourself on the back for healthy thoughts and cognitive strategies. Fast feedback loops are key to fast learning. I like to incorporate the obvious physical motion for positive reinforcements: curling my fingers into a thumbs-up when I count them.

TAP Review

Set a Yoda Timer to review all the TAPs you've tried to install in the last month and figure out what works for you.

Daily Challenge

Have you ever hit the FACTORY RESET button? Share an experience about finally dropping a long-term project, long-held belief, or long-loved identity.

Design 2

This is part 14 of 30 of Hammertime. Click here for the intro.

> *I am a finger pointing to the moon. Don't look at me; look at the moon.*

Rationalists drone on and on about how our fake our models are, how we gesture at and point to deep inarticulate truths, and – to shoulder some of the blame – the importance of circumambulating the truth rather than honing in on it directly. We spend all too much time insisting we're fingers pointing at the moon.

Hammertime says: Fuck the moon.

There are trillions of indistinguishable giant space rocks floating around in the universe. But a human finger contains a trillion copies of the source code for the most power intelligence to walk the known universe. If I had to choose, I'd rather spend my days studying fingers than moons, and it's not even close.

Hammertime is a set of fingers pointing at the moon. Occasionally, it may prove useful to sit back, cross your eyes, and look for the moon: that grand overarching cognitive strategy behind these techniques. But if you miss the moon, fingers are awesome too. So don't worry. Relax. Just do exactly as I say.

Day 14: Design

Previously: Day 4.

Design is the practice of seeing all the tiny incentive gradients in the environment, and shifting them in your favor. Last time, we took environment to mean physical space, but Design principles apply across domains.

Today I will apply Design principles to the design of Schedules, Social Groups, and Screen Space. As budding self-help guru I dub these (together with Space) the Four S's of Design.

Keep in mind the three principles of Design:

1. Intentionality: notice all the knobs you can turn. Turn them the way you intend.
2. Amortization: pay up-front costs to save attention in the long run.
3. Reflexive Towel Theory: the aesthetics of your environment determine your self-image.

Schedules

I am no expert on using calendars; this section is about the basics.

What's the single most important incentive gradient to fix about a calendar? The incentive to use it at all.

Knowing where you'll be, what you'll be doing, how much of your project will be done days, weeks, and months in advance is great. Unbelievably great. It would seem as if

the incentives are already there. So why don't people plan everything all the time?

Everyone has different aversions, but I think the biggest one against calendars is categorizing them as *productivity tools*. My emotions when I first started filling in tasks on the page were those of an unwilling serf hauling his fall harvest to the landowner. That despot wanted my time, all of it, to grind into "productivity." He would give me nothing in return.

Open your calendar now. It is just a tool. Whatever it is that you really want, it's here to help you achieve it. If you truly want to produce productivity, block that off on your calendar. But if you want to binge-watch Death Note this weekend, block that off. If you want guilt-free evenings to lay in bed and cry, block those off too. And treat your calendar reminders as the gentle urging of a well-meaning friend.

Never let your calendar become your tyrant.

Exercise: set a Yoda Timer to plan as densely and as far into the future as you can.

Social Groups

Jordan Peterson likes to say that in the evolution of *homo sapiens*, the Nature that selects in natural selection is three parts natural environment and seven parts other human beings. For the last million years, social, and especially sexual, pressures far outweighed the pressures of survival. The social environment is for us as unchanging and unyielding as the Antarctic winter, and it's hidden incentive slopes have been shaping our lives since millions of years before we were born.

You have the power to shape your social incentives. Reinforcement learning is the primary mechanism by which human beings learn, and we receive so much of our feedback from the social environment, so engineering your social feedback loops is vitally important.

Rule Three in Twelve Rules for Life is: make friends with people who want the best for you. Not everyone shares your values. Not everyone who does can recognize your progress. Not everyone who recognizes knows how to reward. Make friends who reward you for your virtues and punish you for your vices. Ask your friends to hold you accountable, and receive feedback warmly.

Nothing heals the soul like a good smacking from a close friend.

Exercise: set a Yoda Timer to engineer your social environment. Perhaps you want to install a TAP for thanking people for good advice. Perhaps you can teach by example and praise the good you see in people. Perhaps you need to show people you can take criticism. Perhaps you simply need more and better friends.

Screen Space

A mathematician is bound to make a fool of himself teaching macros and keyboard shortcuts to an audience of mostly programmers, but every so often I run into the odd Windows programmer who doesn't use AdBlock. This post is for you.

I have two general principles for the Design of my experience on the computer.

First, never do with a mouse what you can accomplish more efficiently with a keyboard. There are keyboard shortcuts for everything. Set a Yoda Timer in Chrome by typing “Ctrl-T timer 5 minutes.” Archive selected emails with the e key. Jump back to Today in Calendar with the t key. Did I mentioned vim?

Second, build *gentle* incentive slopes. Remove Netflix from your bookmarks to push it one more click away. Set your LaTeX editor to run on startup to make it slightly easier to start writing your next paper. Take full advantage of the taskbar to visibly place the applications you value most.

Things you didn’t know you needed: LyX, vim, AdBlock, HoverZoom, RES, RSS Reader, EXTRA MONITORS.

Exercise: set a Yoda Timer to optimize your Screen Space. Keyboard shortcuts you’d like to practice. Aliases you need to set. Icons you’d like to move around. Look for and eradicate all repetitive actions. There’s a little thing called a computer designed to do that for you.

Daily Challenge

Contribute your vastly superior knowledge of computers to the Design of Screen Space.

CoZE 2

This is part 15 of 30 of Hammertime. Click here for the intro.

Another of CFAR's running themes is: Try Things!

> When you're considering adopting new habits or ideas, there's no better way to gather data than *actually trying [...]* This is particularly important because when something *does* work out, *you get to keep doing it*.

Hammertime will suggest lots of object-level advice. Try them all! A one-in-ten success rate may not feel encouraging, but you can repeat anything that actually works hundreds or thousands of times throughout your life.

Here's a rule of thumb: if there's a 1% chance it'll regularly help in the long run, it's worth trying for five minutes.

Day 15: CoZE

Previously: Day 5.

The basic CoZE experiment technique is:

1. Pick an experience to explore. This should be outside your comfort zone.
2. Devise an experiment or series of experiments. Deconstruct your path from Point A to Point B into palatable baby steps.
3. Try it! At each step, pay close attention to your internal experience, and make sure you're not forcing yourself into anything. You're free to stop at any point.

Today I dispel the illusion that every CoZE experiment should be glamorous. Then, I integrate Aversion Factoring directly into the technique.

Unglamorous CoZE

When I first learned about CoZE, I immediately imagined awesome, courageous, and glamorous experiments. Breaking through to my deepest emotions after subsisting for a month on nootropics and Buddhism, while stranded naked in Siberia. Lucid dreaming in a group hug with Kalahari bushmen while skydiving. Doing a one-finger handstand balanced on a unicycle while delivering extemporaneous limericks to Carnegie Hall.

Your comfort zone limits you in all directions, not just the glamorous ones. The most useful direction to expand can be orthogonal or even opposite to the instinctively shiny ones.

Unglamorous CoZE is expanding in these directions. Breaking down private fears and aversions that nobody will congratulate you for conquering. Trying out socially discouraged activities and points of view. Expansion towards an unappealing role doesn't mean you have to inhabit that role forever – it just gives you a peek into your own versatility, the multitude of roles you could inhabit in different circumstances.

Exercise: Pick a glamorous CoZE experiment you tried in the past. Design a CoZE experiment to grow in the reverse direction. Set a Yoda Timer and explore it!

Aversion Factoring and the CoZE Recursion

Previously: Day 7.

It's high time we start building compound exercises out of our Hammertime techniques. Aversion Factoring fits seamlessly into the prep work for a CoZE experiment. Last time on CoZE, we refrained from attempting CoZE in directions blocked by noticeable aversions, but with the help of Aversion Factoring, we're ready to tackle these tougher challenges.

Recall the three steps of Aversion Factoring:

1. Articulate Aversions: List as many aversions as you can. Be honest and pay attention to trivial inconveniences.
2. Decide Whether to Endorse: Determine if each aversion serves a valid purpose.
3. Solve or Reduce: Try to modify the activity to solve endorsed aversions. Use CoZE to wipe out unendorsed ones.

This brings us to our first compound hammer: the CoZE Recursion.

The CoZE Recursion

1. Pick an experience to explore.
2. Devise an experiment or series of experiments.
3. Aversion factor each step: Articulate your aversions. Modify the experiments to minimize endorsed aversions. Recursively apply CoZE to wipe out un-endorsed aversions.
4. Try the modified experiment(s).

Example:

CoZE public speaking. Notice aversion to all social situations. CoZE speaking with individuals. Notice social aversion due to (endorsed) insecurity about fashion sense. CoZE clothes-shopping. Notice aversion to frivolous expenditure.

God help you if the last aversion had expanded into an infinite loop: *Notice aversion to buying clothes because lack friends with good fashion sense. CoZE making friends...*

You may discover aversions during the experiments themselves. This is fine. Continue to Aversion Factor them. Difficult bugs can generally require up to three layers of recursion.

Exercise: Pick a moderately scary (4-7 on the Bug List) experience to CoZE on up to. Set a Yoda Timer to design CoZE experiments to make progress towards it. Find a time to execute them in the near future.

Daily Challenge

Today's challenge is a question: is courage just the absence of fear?

If there is a meaningful difference between the two, is CoZE primarily about increasing courage or reducing fear? Whichever it is, is there an alternative method to do the other?

Three Miniatures

This is part 16 of 30 of Hammertime. Click here for the intro.

The sixth day always marks the boundary between concrete and abstract. Today, I mark the occasion with three essays on new techniques.

These essays are short because I lack data and examples. All concept handles and perspectives are preliminary. The two latter essays are, I think, two fingers pointing at the same moon.

Day 16: Three Miniatures

Pressure Points

I need to sleep earlier.

I can't sleep now because I need to get this paper written by tomorrow.

I'll just finish it in the morning.

I don't trust myself to work in the morning.

I need to try harder to trust myself and sleep earlier

This train of thought plagued me in a past life. Do you see what I was missing?

I was pushing on the wrong side.

I always lazed about in the morning. In this scenario, "Try harder to trust myself" is self-delusion. To fix it, I first needed to cultivate a habit of working in the morning, or at least being able to. Once I could do that, the original train of thought would automatically cut itself off in the middle.

Getting work done in the morning wasn't easy, but it was the right place to try really hard. Once I solved that problem, I could trust my next-morning self. It became easy to correct my sleep schedule.

Pressure Points is a lateral thinking technique. For any given problem, there are many places to apply pressure, and all it takes is to find the pressure point to apply brute force most efficiently. The pressure point is rarely the obvious direction: chances are, you've already been pushing in that direction, and it hasn't helped thus far. Look for counter-intuitive places to apply brute force.

Here are three examples of Pressure Points and its particular brand of creativity:

Lucid Dreaming is all about finding the right Pressure Points. Instead of "Intend really hard to lucid dream," the two main techniques are practicing reality checks while awake, and keeping a dream journal to improve dream recall.

People often approach social anxiety with “Try to care less about what other people think.” This is about as effective as “Try not to notice your breathing.” A Pressure Point approach to social anxiety is “Try to focus on other people’s body language and notice their anxiety.”

I’ve been working on TAPs for building better posture. The only one that has had any effect is “Turn to face the shower nozzle.” When I face away from the shower nozzle, I hunch over to avoid the feeling of water on the back of my neck. When I face towards it, I pull my head back and puff my chest out to get my face out of the spray.

History Search

I need to make a confession.

I’ve been cheating at Hammertime.

Half of the personal examples for any given technique come from before I even knew about the technique. Many of the tentative techniques and variations I propose are more like “patterns I noticed in my past” than the product of consciously design.

Rationality is systematized winning. The thing is, I got good at certain things before I met rationality. We all did. We’ve all been discovering ad-hoc versions of rationality techniques since before the Sequences were ever written.

Each time you learn a new rationality technique, look into your past for all the times you’ve already been doing it. You’ll get a concrete understanding of the technique and feel more ownership over it, and this will also help you adjust your seat to tailor it to your own needs.

Similarly, new rationality techniques can be discovered by searching yours and others’ pasts. Notice the cognitive strategies your brain employs already, and try to articulate them. Remember that articulating unspoken rules is participating in the divine act of creation (see Logos).

Superstition Culling

A man takes a nootropics cocktail called BrainHammer for 30 days. He feels energized and clear-minded, sleeps two fewer hours a day, and gains control over his anger problem. He develops a rosy halo around the drug, and keeps taking it ad infinitum.

BrainHammer is actually ten different drugs, and caffeine is the only one with positive effects. But BrainHammer is forty times the price of coffee, and several of its ingredients come in miniscule, irrelevant doses. One of the acting ones reduces the man’s sex drive, while another ingredient is slowly giving him kidney stones.

Hammertime (and CFAR) can be like this nootropics cocktail. Thirty days later, you come away with a glow of satisfaction, equipped with ten max-level hammers for tackling your toughest bugs. You start pounding away.

It turns out only one of the hammers (Yoda Timers) does all the work. You’re just really motivated by timers and deadlines. Meanwhile, 80% of your Hammertime-inspired rationality practice consists of placebo motions: rearranging furniture, learning three

different yogas and five different meditations, muttering incoherently under your breath, stacking up half-finished journals and spreadsheets, ordering junk and garbage off Amazon. Also, you're doing Internal Double Crux completely wrong and it's slowly making you manic-depressive. You won't notice until it's too late.

Superstitions crop up inevitably whenever happy things happen. It takes discipline and the scientific method to hone in on the active ingredients of a drug cocktail, and the same holds for rationality techniques. If you're learning more than one technique at a time, actively plan to cull out superstitions.

Even if you're learning a single technique with multiple steps, only one of them might be load-bearing. For instance, in high school I learned that the sole value of note-taking is writing down names to remember them. Culling out the superstition, I continued to take notes but stopped keeping them.

Daily Challenge

Set a Yoda Timer and search your history for all the times you improved rapidly. Can you articulate a new rationality technique from those experiences?

Focusing

This is part 17 of 30 of Hammertime. Click here for the intro.

> You know how they say we only use 10 percent of our brains? I think we only use 10 percent of our hearts.

~ Owen Wilson

It is with some trepidation that I venture into the "fuzzy System 1" side of instrumental rationality. I worry that these introspective techniques optimize too much for cathartic eureka moments, and that the resulting feelings far overstate their true value.

Nevertheless, there is a definite power to these methods. You have subconscious beliefs, values, and strategies that you're unaware of, or at least can't articulate. Gendlin's Focusing is a starting point for plumbing these hidden depths.

Day 17: Focusing

Background: "Focusing" for skeptics.

tl;dr: your brain hallucinates sensory experiences that have no correspondence to reality. Noticing and articulating these "felt senses" gives you access to the deep wisdom of your soul.

I'll start by explaining my most gears-like model for why focusing works, and then describe some exercises towards strengthening the Focusing muscle.

One of the predictions of my model is that felt senses are only one piece of the nonverbal puzzle – the patterns in our dreams and our tastes for fiction and mythology, for instance, serve the same function. This will be the content of a future post.

Left and Right Brain

This model is derived from Jordan Peterson's lectures on psychology, and in particular this conversation. I reserve the right to call everything fake if you try to falsify it.

Human beings are both predator and prey. This duality is so central to human evolution that the brain is divided left and right to serve the two different purposes separately.

The left brain is the predator brain, the center for "approach" mechanisms. It's built for tracking a particular prey animal, articulating rules about behavior, and solving concrete problems. To fix your attention on a target is to activate your left brain and get ready to hunt it down. In the direction you look, there is clarity and legibility. Over that direction, you gain power and mastery.

“Sin” derives from the Greek word for *missing the mark*: human beings are aiming creatures.

The right brain is the prey brain, the center for “flight” mechanisms. It’s built for hypothesizing a venomous fog of worst-case scenarios: snakes in every tree, traps under every bramble. The right brain is constantly on edge, searching for subtle clues of being tracked by a clever predator or failure mode. It operates on the things you don’t know and cannot see: the space behind your head, the shadows in dark corners, the places and concepts you circumambulate.

With its higher level of clarity and certainty, the left brain is by far the more verbal of the two, and most of your articulated knowledge resides there. The right brain, on the other hand, may have access to the most important big-picture insights about your life. The trouble is to communicate them.

When the right brain has a message to send that won’t go directly through the corpus callosum, the message manifests in other ways. You feel a tightness in your chest or a glow in your belly. Unbidden images appear to you when you close your eyes. Recurring nightmares play out the last moments of your likely doom.

Focusing is about noticing these subtle clues and completing the communication between left and right brain.

Felt Senses

The basic idea of Focusing is to notice and track your felt senses and learn to articulate them. The most exciting thing that happens during focusing is noticing a “felt shift,” a relief or change, in the sensation once you hit upon the right words to frame it. This response is your right brain confirming that you got the message.

I’ll start by listing a few felt senses I’ve had recently:

- When I solve a problem in a creative way (e.g. fix posture by turning in the shower), there’s a sensation of enlightenment at the back of my head which literally feels like my skull is opening up. The words to this feeling are “I’ve discovered a new dimension!”
- I sometimes sit slouched over in bed for hours at a time browsing Facebook or Reddit, playing video games, or binge-watch a season of a TV show. After getting up from the slouch, my whole body is enveloped in a haze of laziness and decay. The zombie haze is thickest inside my ribs. The words to this pressure are “Symptoms of the spreading corruption.”
- A piece of my social anxiety forms a hard barrier that pushes against the center of my chest. I learned the words to this feeling from a post by Zvi: “Conform! Every time you walk outside the norm, think about the implicit accusation you’re making against everyone who didn’t try it.”

Here’s Gendlin’s Focusing check from CFAR:

1. Say aloud “Everything in my life is fine,” or “I’m on track with all of my goals.”

2. Pay attention to the sensations in your belly, chest, and throat. If you’re like most people, something will catch or react weirdly to the statement.

3. Try to get a sense of what the feeling “sees,” and write it down.

4. Imagine setting that thing aside (like putting it next to you on a park bench), and try again: “Apart from that, everything in my life is fine.” See what catches this time.

5. Continue until you reach a statement that doesn’t produce a reaction, and instead *rings true* (e.g. “Apart from A, B, C, D, my life is fine right now.”)

Set a Yoda Timer and try the Focusing check.

Daily Challenge

Share a felt sense and its True Name.

Goal Factoring

This is part 18 of 30 of Hammertime. Click here for the intro.

Up until today, Hammertime focused on improving one's ability to achieve one's goals. The next two techniques, Goal Factoring and Internal Double Crux, are designed to figure out what goals to pursue. For the largest goals in life, you should be able to make a detached decision about whether they're worth pursuing before you throw your all into them.

Day 18: Goal Factoring

Previously: Day 7, Aversion Factoring.

Goal Factoring is a CFAR technique for systematically figuring out all the subgoals and aversions you have around an action, and what to do about them. The basic algorithm:

1. Pick an action. It can be something you already do.
2. Factor the action into goals and aversions. Write down all the costs and aversions to pursuing the action, and continue to factor sub-goals until they feel like irreducible components.
3. Brainstorm possible replacement actions. Try to design another action that achieves the goals better and reduces the costs and aversions. This action can be an upgrade of your current action, something else altogether, or even a combination of two or more actions. Make a new plan.
4. Reality check. Imagine instituting your new plan. Decide if you're satisfied. Also, decide if it's feasible by Murphyjitsu.

This is already quite a complicated and useful beast. Three things to keep in mind:

Use a focusing check to find all the subgoals and aversions. If I say out loud, "The only reason I want to go to the gym is physical health," I feel a curtain of discontent that reminds me physical attractiveness is also important. Remember that honesty and attention to detail are essential to finding aversions, and this applies to goals as well!

Goal factoring might solve the problem at any step. Writing down your true motivations can be enough to figure out the right course of action. About three months ago, I noticed that the main motivation for my video game addiction is "Prove to my parents that it's possible to be successful without giving up video games." Writing this down made it impossible to endorse this action any longer.

Prepare to accept all possible worlds. Keep an open mind going into Goal Factoring: you're allowed to consider all the alternatives. You're also allowed to keep doing what you're currently doing afterwards. Try to release any attachment to the action itself beyond its instrumental value. Get a little worried if your main reason to do action X is to become the kind of person who does X, but at least write this down as an explicit sub-goal.

Exercise: pick an action or habit you want to pick up or drop, and set a Yoda Timer for 20 minutes to Goal Factor it.

Daily Challenge

Set a Yoda Timer to Goal Factor “do Hammertime.” Share your motivations and aversions.

TDT for Humans

This is part 19 of 30 of Hammertime. Click here for the intro.

As is Hammertime tradition, I'm making a slight change of plans right around the scheduled time for Planning. My excuse this time:

Several commenters pointed out serious gaps in my knowledge of Focusing. I will postpone Internal Double Crux, an advanced form of Focusing, to the next cycle. Instead, we will have two more posts on making and executing long-term plans.

Day 19: TDT for Humans

Previously on planning: Day 8, Day 9, Day 10.

Today I'd like to describe two orders of approximation to a working decision theory for humans.

TDT 101

Background reading: How I Lost 100 Pounds Using TDT.

> Choose as though controlling the logical output of the abstract computation you implement, including the output of all other instantiations and simulations of that computation.
>
> ~ Eliezer

In other words, every time you make a decision, pre-commit to making the same decision in all conceptually similar situations in the future.

The striking value of TDT is: make each decision *as if you would immediately reap the long-term rewards of making that same decision repeatedly*. And if it turns out you're an updateless agent, this actually works! You actually lose 100 pounds by making one decision.

I encourage readers who have not tried to live by TDT to stop here and try it out for a week.

TDT 201

There are a number of serious differences between timeless agents and human beings, so applying TDT as stated above requires an unacceptable (to me) level of self-deception. My second order of approximation is to offer a practical and weak version of TDT based on the Solitaire Principle and Magic Brain Juice.

Three objections to applying TDT in real life:

Spirits

A human is about halfway between “one monolithic codebase” and “a loose confederation of spirits running a random serial dictatorship.” Roughly speaking, each spirit is the piece of you built to satisfy one primordial need: hunger, friendship, curiosity, justice. At any given time, only one or two of these spirits is present and making decisions. As such, even if each individual spirit as updateless and deterministic, you don’t get to make decisions for all the spirits currently inactive. You don’t have as much control over the other spirits as you would like.

Different spirits have access to different data and beliefs. I’ve mentioned, for example, that I have different personalities speaking Chinese and English. You can ask me what my favorite food is in English, and I’ll say dumplings, but the true answer 饺子 feels qualitatively better than dumplings by a wide margin.

Different spirits have different values. I have two friends who reliably provoke my “sadistic dick-measuring asshole” spirit. If human beings really have utility functions this spirit has negative signs in front of the terms for other people. It’s uncharacteristically happy to engage in negative-sum games.

It’s almost impossible to predict when spirits will manifest. Recently, I was on a 13-hour flight back from China. I started marathoning Game of Thrones after exhausting the comedy section, and a full season of Cersei Lannister left me in “sadistic asshole” mode for a full day afterwards. If Hainan Airlines had stocked more comedy movies this might not have occurred.

Spirits can lay dormant for months or years. Meeting up with high school friends this December, I fell into old roles and received effortless access to a large swathe of faded memories.

Conceptual Gerrymandering

Background reading: conceptual gerrymandering.

> I can make a problem look either big or small by drawing either a big or small conceptual boundary around it, then identifying my problem with the conceptual boundary I’ve drawn.

TDT runs on an ambiguous “conceptual similarity” clause: you pre-commit to making the same decision in conceptually similar situations. Unfortunately, you will be prone to motivated reasoning and conceptual gerrymandering to get out of timeless pre-commitments made in the past.

This problem can be reduced but not solved by clearly stating boundaries. Life is too high-dimensional to even figure out what variables to care about, let alone where to draw the line for each of them. What information becomes salient is a function of your attention and noticing skills as much as of reality itself. These days, it’s almost a routine experience to read an article that sufficiently alters my capacities for attention as to render situations I would previously have considered “conceptually similar” altogether distinct.

Magic Brain Juice

Background reading: Magic Brain Juice.

> Every action you take is accompanied by an unintentional self-modification.

The human brain is finicky code that self-modifies every time it takes an action. The situation is even worse than this: your actions can shift your very values in surprising and illegible ways. This bug is an inherent contradiction to applying TDT as a human.

Self-modification happens in multiple ways. When I wrote Magic Brain Juice, I was referring to the immediate strengthening of neural pathways that are activated, and the corresponding decay through time of all pathways not activated. But other things happen under the hood. You get attached to a certain identity. You get sucked into the nearest attractor in the social web. And also:

Exposure therapy is a powerful and indiscriminate tool. You can reduce any aversion to almost zero just by voluntarily confronting it repeatedly. But you have fears and aversions in every direction!

Every move you make is exposure therapy in that direction.

That's right.

Every voluntary decision nudges your comfort zone in that direction, squashing aversions (endorsed or otherwise) in its path.

Oops!

Solutions

I hope I've convinced you that the human brain is sufficiently broken that our intuition about "updateless source code" don't apply, and trying to make decisions from TDT will be harder (and may have serious unintended side effects) as a result. What can be done?

First, I think it's worth directly investing in TDT-like behaviors. Make conscious decisions to reinforce the spirits that are amenable to making and keeping pre-commitments. Make more legible decisions and clearly state conceptual boundaries. Explore virtue ethics or deontology. Zvi's blog is a good a place.

In the same vein, practice predicting your future behavior. If you can become your own Omega, problems you face start looking Newcomb-like. Then you'll be forced to give up CDT and the failures it entail.

Second, I once proposed a model called the "Ten Percent Shift":

> The Ten Percent Shift is a thought experiment I've successfully pushed to System 1 that helps build long-term habits like blogging every day. It makes the assumption that each time you make a choice, it gets 10% easier.
>
> Suppose there is a habit you want to build such as going to the gym. You've drawn the pentagrams, sprinkled the pixie dust, and done the proper rituals to decide that the benefits clearly outweigh the costs and there's no superior alternatives. Nevertheless, the effort to make yourself go every day seems insurmountable.
>
> You spend 100 units of willpower dragging yourself there on Day 1. Now, notice that you have magic brain juice on your side. On Day 2, it gets a little bit easier.

> You spend 90 units. On Day 3, it only costs 80.
>
> A bit of math and a lot of magic brain juice later, you spend 500 units of willpower in the first 10 days, and the habit is free for the rest of time.

The exact number is irrelevant, but I stand by this model as the proper weakening of TDT: *act as if each single decision rewards you with 10% of the value of making that same decision indefinitely*. One decision only loses you 10 pounds, and you need to make 10 consecutive decisions before you get to reap the full rewards.

The Ten Percent Shift guards against spirits. Once you make the same decision 10 times in a row, you'll have made it from a wide range of states of mind, and the exact context will have differed in every situation. You'll probably have to convince a majority of spirits to agree with making the decision.

The Ten Percent Shift also guards against conceptual gerrymandering. Having made the same decision from a bunch of different situations, the convex hull of these data points is a 10-dimensional convex region that you can unambiguously stake out as a timeless pre-commitment.

Daily Challenge

This post is extremely tentative and theoretical, so I'll just open up the floor for discussion.

Friendship

This is part 20 of 30 of Hammertime. Click here for the intro.

There's a serious and scary phenomenon which Valentine's recent posts have been touching on: much of *who you are* only exists (or is expressed) in the presence of other people. In the words of Bishop Berkeley, *esse est percipi*: To be is to be perceived. Hammertime will always be an incomplete endeavor unless it is applied to social settings – there are major chunks of the psyche only accessible in such settings.

Up to now, Hammertime has mostly been a set of tools for the individual rationalist in a social vacuum. Today I want to talk the problem of other human beings, and how to go about designing social interactions that are conducive to the practice of instrumental rationality.

Hammertime Day 20: Friendship

Background: The Intelligent Social Web

There's good evidence in biology that the power of the human brain largely evolved to solve ever-complexifying social problems. Much of the heavy cognitive machinery in your head is primarily built for and responds best to social interaction. Brains are extremely good at detecting social threats and anomalies, at regulating implicit status ladders, at reading body language, and at simulating other brains.

This post is a start at the Design of optimal two-person interactions.

Iterated Games

Rationalists spend a lot of time railing against the failings of causal decision theory, and promoting alternatives that solve them. The uncomfortable truth, however, is that you will not make causal decision theorists cooperate on the prisoner's dilemma by throwing tomes of philosophy at them, and many many people are causal decision theorists. Not all hope is lost though: there's a known, albeit unglamorous, solution to coordination failures within the framework of causal decision theory: iterated games.

Iteration is the easiest path to building strong friendship: make interactions longer and more regular.

In the middle of January, I began contacting friends and setting up regularly weekly chats. Almost nobody refused. A handful of interactions fizzled out, but the ones that lasted have been unbelievably positive. I kept ramping up the number of conversations until it felt actively fatiguing. Today this habit alone allows me to talk to an average of one extra person per day for an hour and a half.

Human beings are unbelievably reciprocal creatures in stable long-term relationships. The incentives are quite robust. Jordan Peterson once highlighted this with a pithy phrase about marriage (paraphrased): "You can't win an argument against your wife if she loses. After all, you still have to live with her."

Of course, human beings are also stupid and perverse enough to ignore the strongest incentives. How many millions of life-long partnerships ended with decades of mutual abuse? Keep your eyes open.

Conversation 101

Here are three object-level ideas for having useful conversations.

Socratic Ducking

> Rubber Ducking
>
> Getting a person to act as a rubber duck who you talk your ideas out to in order to get a clear handle on them.
>
> Socratic Ducking
>
> Aiding a partner in thinking through an idea or solving a problem. Combines socratic questioning and rubber ducking. Attempt to offer few suggestions and thoughts while instead alternating between stimulating questions and attentive silence. Encourage the other person to think through complex threads and think deeply about the ramifications of ideas and possible solutions.

Oftentimes there is a clear listener and talker in a conversation. As the listener, focus primarily on paying attentive silence and occasionally asking pointed or clarifying questions when the conversation seems to dry up. The primary goal is to keep your partner generating ideas and on track.

A friend of mine stimulated a major breakthrough in a session of aversion factoring for me by nodding silently the whole time, except for uttering a single well-timed word: “try!” This encouraged me to expend the necessary mental effort to break through that mental barrier and correctly identify an aversion towards planning.

ITTs

> *The Ideological Turing Test is a concept invented by American economist Bryan Caplan to test whether a political or ideological partisan correctly understands the arguments of his or her intellectual adversaries: the partisan is invited to answer questions or write an essay posing as his opposite number. If neutral judges cannot tell the difference between the partisan’s answers and the answers of the opposite number, the candidate is judged to correctly understand the opposing side.*

Intellectual (Ideological?) Turing Tests, or ITT’s, can be rather laborious. The minified conversational norm is: you are not allowed to move forward with an argument until you have accurately summarized the other person’s point of view *to their satisfaction*.

Rabbitholes

Conversations can get derailed rather rapidly, and it’s a well-established fact that all conversations after midnight will devolve into a debate about consciousness.

For online conversations, I make a habit of collecting possible tangents on a sheet of paper when they come to mind, instead of immediately tossing them into the fray and risking the entire current train of thought. There will always be time later for your fascinating point.

Take a Yoda Timer to train the following TAP: whenever a related conversation topic comes to mind, ask yourself whether you want to go down that rabbit-hole.

Daily Challenge

Book a 30 or 60-minute chat with me on Calendly to talk about anything.

[Update: This challenge is still active as of 10/2021 and will be for the foreseeable future. Please take me up on it! The conversations I've had over the years because of this post are among the best side effects for me personally from writing this entire sequence.]

Hammertime Intermission #2

The final cycle of Hammertime will return on 3/12.

This is an open thread for feedback about Hammertime.

Other Resources

Owen Shen has been writing about instrumental rationality for a long time at MindLevelUp. This material is much more coherent and actually cites the science behind the techniques.

SquirrelInHell has taken on a similar project but at a much deeper level at BeWellTuned. These techniques are BLACK MAGIC powerful but may cost your soul. Consider Hammertime a gentle tutorial mission for BeWellTuned.

Epistemic Rationality

A piece of the third cycle may be devoted to applying instrumental rationality tools towards increasing epistemic standards in practice.

What are the immediate applications of Hammertime tools towards truth-seeking? Are there TAPs, Design decisions, CoZE experiments, or Focusing techniques that you would immediately recommend to increase epistemic standards? For example, I've seen people claim that they've installed a gut-level instinct for Bayes Theorem.

How does truth-seeking interface with instrumental rationality, especially on corrupted hardware? We pay a whole lot of lip service to "I desire to believe what is true," and yet an enormous amount of our practical advice sounds like "Act as if [obviously false statement] is true."

Meta-Cognitive Blind Spots

I have a friend who got so excited about conquering new fears that he's been constantly flitting around, looking for aversions to squash with CoZE. When I asked him to try sticking to one thing for a while, he nearly refused, saying "It feels hard to look at." My words:

> when ppl get excited about CoZE
>
> there's something ridiculous about it
>
> you're literally looking for things that scare you
>
> like actually scare you
>
> and that's not fun

One time this week, I was Focusing on the feeling of hating algebraic geometry homework. I found the feeling's true name: "Working on problems in other fields

makes me feel low-status.” The aversion completely vanished, but I was left with the sinking feeling that I’m status-seeking all the way down. In fact, this whole paragraph reporting this character flaw? It’s a self-deprecation status move too. Fuck.

[Commenting guidelines: At least one line of poetry per comment.]

Bug Hunt 3

This is part 21 of 30 in the Hammertime Sequence. Click here for the intro.

I took a long break from Hammertime to check the fundamental question: *am I actually better at achieving my values now?*

The answer is a solid yes. Problems that used to live in the category of "not within my capabilities" disintegrated into so many puffs of malevolent smoke. Paper-writing got itself done. Fifteen hundred words of half-decent fiction got written every day. For the first time in my life, I live in such a thoughtfully decorated room I'll actually miss it when I move away. I felt like a rationality Warlock:

A high-level Aversion appears...Hah! With the power of FOCUSING, I'll scry your true name, Demon!

"Status Regulation, Begone!"

This section feels impossible to write...

I know! I'll do it in FIVE MINUTES!

I have no idea what my problem is...

Fear not! I'll blast it with the magic of FRIENDSHIP!

So you're stuck on a quest to save the world...

Have you tried REMOVING TRIVIAL INCONVENIENCES?

If you're reading Hammertime simply for my scintillating wit, that's completely fine! Just remember that these techniques might also help you achieve your values if you give them a chance.

Hammertime: The Third Cycle

> *Twice and thrice over, as they say, good is it to repeat and review what is good.*
>
> ~ Plato.

The third cycle is ten days of review. On each day, we will attempt to tease out the unifying meta-principles behind each technique, taking them (and all the others) to the limit of their power. Here's a tentative schedule:

1. Bug Hunt 3
2. Yoda Timers 3: Speed
3. TAPs 3: Reductionism
4. Design 3: Intentionality
5. CoZE 3: Empiricism
6. Growth Triplets
7. Internal Double Crux: Duality
8. Focusing 2: Fusion

9. Murphyjitsu 2: Humility
10. TDT 2: Post-Consequentialism

Day 21: Bug Hunt 3

Today we're back to Bug Hunt with three more sets of prompts to help find the biggest bottlenecks in your life. After you read each, set a Yoda Timer to brainstorm bugs.

1. Getting Got

The world is Out to Get You. Social media. Capitalism. Your job. Your family. Your friends. Your hobbies. Everyone wants your time, money, and attention. How do you keep yourself from Getting Got all the time?

Do you know how to say no? If you don't Get Gone regularly, you're easy pickings. Things are often worse than they appear. Things deteriorate over time. Things want more and more of your soul. There's no such thing as a one-hour game of Civ. Get Gone. You don't owe anybody everything.

Do you know how to set boundaries? Some things can only be Worth It if you can draw a line in the sand. Set a budget. Or a timer. Get Compact and hold the line as if your life depends on it.

2. Hamming Problems

Background reading: Anxious Underconfidence and Status Regulation

> What are the important problems of your field?
>
> What important problems are you working on?
>
> If what you are doing is not important, and if you don't think it is going to lead to something important, why are you … working on it?

~ Richard Hamming.

To take an incremental approach: are there slightly more important problems you could be working on? Why aren't you working on them?

Anxious Underconfidence is an artifact of an ancestral environment where every failure is fatal. Do you have Anxious Underconfidence? How often have you failed on a significant undertaking in the last year? Don't maximize your percentage of wins. Maximize total number of wins. That's what counts.

Do you use status as a proxy for competence? Do you believe that only people with tenure, wealth, age, or social capital have the right to work on important problems? Is your assessment of your own abilities a function of how others perceive you?

3. Fail Gracefully, Succeed with Abandon

Background reading: Failing with Abandon.

There's a Chinese idiom, 破罐子破摔, which means: "might as well smash a cracked pot." Failing with abandon is angrily smashing a pot with the slightest crack. "I didn't like it anyway!"

Does that appeal to you?

Failing with Abandon ignores the fact that utility functions are usually continuous. Failing a little is OK. Keep at it. Something is better than nothing.

Failing with Abandon takes away valuable learning experiences. If the last homework can't save your grade, do you still put in the same effort? If you're twenty points down in a game of Go, do you still try your best? Or do you just go through the motions? Life is a very long iterated game, and Failing with Abandon is forfeiting the future.

On the flip side, do you always satisfice? Do you turn in the bare minimum to make a GPA? When you hit the target, do you run home to party? If you're up twenty points in a game of Go, do you play improper but safe moves to secure the win? Satisficing is giving up an opportunity to reach your full potential.

Failing with Abandon and satisficing are both symptoms of near-sighted hyperbolic discounting.

Instead, fail gracefully. Succeed with abandon.

Daily Challenge

Are you better at achieving your values since Hammertime Day 1? If so, what helped?

Yoda Timers 3: Speed

This is part 22 of 30 in the Hammertime Sequence. Click here for the intro.

At some point around the end of high school, being *fast* became unfashionable. When did this happen?

Why do we channel so much more energy into doing more difficult things, instead of doing simple things faster? How much faster could you do your job? Two times faster? Five times?

Instead of *I want to be stronger*, say *I want to be faster*.

If you pay attention to speed, you might just find a way to do a whole week's worth of work in five minutes.

Day 22: Speed

Here are three exercises to help you find that rush of *mind working on overdrive*:

1. Typeracer. Play this game for five minutes. How much faster did you get at typing? What did you learn?
2. Go on YouTube and watch a talk at 2x speed for five minutes. If you're having trouble following, turn on subtitles. Notice that this is possible.
3. The Arithmetic Game. Play three rounds of this with standard settings. How much faster did you get?

Here are three principles I've extracted from using Yoda Timers to do everything faster.

1. Mistakes Are Fatal

When I first played Typeracer, I started out at a measly 70 wpm and worked my way up to around 90 by just trying harder. Eventually, I hit a plateau because I was still constantly making typos and backspacing. Each mistake cost the time of four or five characters. The backspace key was my Achilles heel.

That's why I forced myself to slow down and get everything right the first time. At first, this lowered my wpm, but with a bit of work, my fingers felt more nimble and intentional. I cut down the number of typos I made by a factor of 4 or so – it turns out there's a handful of sequences of keys I constantly get wrong or out of order. My wpm skyrocketed to 120.

In real life, mistakes are even costlier. Getting sick is way costlier than having good hygiene. In programming, it's common knowledge that testing and bug-fixing takes at least three times as long as writing code in the first place. In math, months of paper-writing can go down the drain when you finally notice a severe and unfix-able logical misstep. At the Olympics, every single mistake will cost you the medal.

If you want to be faster, you must have zero tolerance for even the slightest errors, and slowing down (at first) to practice perfectionism will be worth it. Get it right the first time.

2. Speed Limits are in the Mind

Everyone has a rough idea of how long things have to take. Solving a hard research problem always takes at least a month, right? Writing a paper should take at least an hour, right?

When I first started playing the Arithmetic Game for middle school MathCounts training, my high score was close to 20. After a few months of dedicated training, my record hit 90, making it onto the leader-board of the time.

For any given task, do not assume you're doing it anywhere close to your real speed limit. It used to take me at least four hours to write a blog post this length. This one clocks out in just under forty minutes.

3. Speed is Easier than Strength

In intellectual work, it's *much much* easier to get twice as fast than twice as good. It's much easier to multiply twice as quickly than to learn to solve harder problems. It's much easier to type twice as much content than to write twice as well.

Human beings are supremely good at training rote tasks to maximum efficiency. Take advantage of that. Learn to read twice as fast, write twice as fast, talk twice as fast, walk twice as fast, watch videos twice as fast. I've been watching videos at 2x speed for as long as I can remember, and I can't even tolerate regular speed anymore. Once you habituate to going faster, you reap all this free energy that was just lying around while you were waiting.

Speed is underrated. Short training sessions focused on speed will create lasting impacts on your productivity.

Daily Challenge

Share your proudest speed record. Fast is fashionable again!

TAPs 3: Reductionism

This is part 23 of 30 in the Hammertime Sequence. Click here for the intro.

In school, we spend thousands of hours learning about the building blocks of the universe. We learn that reality reduces into little pieces: organisms into cells, books into pages, skyscrapers into atoms.

Your life belongs inside this infinitely divisible reality. Your psyche divides into subpersonalities, emotions into qualia, actions into goals and aversions, habits into TAPs. In fact, what we think of as objects are usually *patterns of interaction* between many tiny pieces.

Day 23: Reductionism

Trigger-action plans are the building blocks of habits – all habits can be built out of single steps.

I want to share a model for why it's so important to break actions down with reductionism.

Zeno's Paradox Retold

Here's the old paradox of Zeno:

To win a race, you have to run the first half. Before you finish the first half, you must complete the first quarter. Before you finish the first quarter, there's the first eighth, and so on ad infinitum. Thus, by halving the first segment, every race is divisible into infinitely parts, and to complete the race you must make infinitely many actions.

What can we learn from Zeno's paradox?

Of the infinitely many steps in the race, the first step accomplishes almost all of them. It follows that the first step in a race is infinitely more difficult than every later one.

The Method of Exhaustion

From Zeno's Paradox, we readily derive the following algorithm for deconstructing problems:

1. Pick an action.
2. Divide it into halves. Focus on the first half.
3. Repeat to exhaustion.

For example, I can decompose the action of "write a blog post" in exponentially ascending order of difficulty:

1. Take a deep breath.
2. Visualize success.

3. Turn on computer.
4. Open Chrome.
5. Log in.
6. Type a letter.
7. Type a word.
8. Type a sentence.
9. Type a paragraph.
10. Type a section.
11. Type a post.
12. Click “publish.”

After having completed the method of exhaustion, executing the action is much easier. Notice that even though I’m ostensibly only 1/3 of the way through writing this post, I’ve already accomplished 10.5/12 steps in the workflow.

I’m almost done!

Steps of Equal Difficulty

You may think the last section was flippant or self-delusion.

Nope.

I’m completely serious.

Walk through the whole activity of blogging (if blogging’s not aversive to you, pick whatever else you’re procrastinating on and apply the method of exhaustion to that one instead), and note how much total mental resistance you push through at each step in the 12-step process. Also note how likely you are to give up at each step.

The normal method of planning is to break into equally sized blocks, where size means “time and effort in objective reality.” Take stock of all the plans you’ve made in your life. How many failed at the very beginning? How many failed near the middle? How many failed towards the very end?

Most things fail before they begin. Of the ones that do begin, most fail immediately.

You don’t live in objective reality. You live in the mad world of Zeno, where the first step is infinitely difficult. The Method of Exhaustion is designed to parse a hard problem into steps of roughly equal *psychological* difficulty and failure rate.

Exercise: Apply the Method of Exhaustion to your next big project. How many pieces did you break it into?

Daily Challenge

Share anecdotes or data on how long it takes [intentions, projects, plans, relationships, careers, startups] to fail. What do the curves look like?

Design 3: Intentionality

This is part 24 of 30 in the Hammertime Sequence. Click here for the intro.

Intentions are momentary, but problems last forever.

A human being's attention flits around like the Roman God Mercury, root of the word "mercurial" – *subject to sudden or unpredictable changes of mood or mind.* The biggest problems in life require concentrated effort over years or decades, but you can only muster the willpower to even intend to solve a problem for minutes or hours. Worse, you can pretty much only maintain one intention at a time.

How do we make intentions count?

The philosophy of Design is: *build intentions into external reality*. Like your problems, external reality also lasts forever.

Day 24: Intentionality

You need to lose those love handles. Your reading list is piling up. You need to learn ten different programming languages. You need to sleep three hours earlier. You need to maintain your closest friendships. You juggle three different addictions that take turns monopolizing your life. You need to present like a functioning adult to your parents and coworkers. A childhood trauma you're repressing makes it impossible to befriend a certain half of the population.

You have a lot of problems, each of which requires dedicated effort and thought to fix. Worse, each problems deteriorates while you're working on the others. Perhaps some have gone so neglected they're impossible to look at, and are slowly swallowing the rest of your life like a super-massive black hole.

Right this minute, there's probably only a handful of problems that feel alive enough to you to inject energy into. Of those, you can only work on one at a time. In this crazy unfair world, how do you make the most of your intentions?

Outsource the Burden

There's a certain unproductive way of thinking which goes like this:

"If I were really rational, I wouldn't need all these aids. I wouldn't need chrome extensions to block Facebook and Twitter, friends to reward me for the slightest progress, and SSRI's to keep my demons at bay. I could just do what is right all the time."

Give it up. There might be something aesthetically appealing about handicapping yourself this way, but it's no way to actually solve problems. Life is tough and deeply unfair and you'll need all the help you can get if you want any chance of success.

Part of the Design philosophy is allowing yourself to outsource your heroic burden. You can't complete this quest alone. Make all the inanimate and animate objects in your

life sidekicks in your quest – not obstacles. Every tiny push in the right direction you can get externally is one less ounce of force you need to generate yourself.

Incentive Gradients

The world is filled with tiny incentive gradients that slowly push you towards local optima. Look for and pay attention to these incentive gradients so that you can turn them to your advantage. Tipping the scales in the smallest way can do work for you in the long term.

In practice, we focus on the 4 S's of Design. All of these we've already covered, but it's time to review again.

Space. How is your space designed to help achieve your goals? Is the place you work maximally comfortable and well-lit? Are the things you need for your routines placed in optimal locations? Does the aesthetics of the space properly reflect your values? Is it conducive to productive social interaction?

Schedules. How do you manage your time and energy throughout days and weeks? Do you work better by interleaving different kinds of activities, or by batching? Do you schedule things in such a way that you look forward to the future? Do you use Calendars and apps efficiently to remove the mental load of remembering things? Do you follow your plans?

Social Groups. Do your friends reward you for making progress? Do they punish you for failure? In any social network, every individual is drawn inevitably into a niche: the Silent One, the Alpha, the Clown, the Cheerleader, the Cynic. What niche do you inhabit? What forces push you there? Is it where you want to be?

Screens. Given how much time we spend on screens, and the Machiavellian motions by which everything on the internet tries to ensnare your soul, pay attention to your computer habits. Draw a quick graph of how you navigate applications and websites. What factors take you from one place to the next? Where do you get sidetracked most often?

Be Good Incentives for Others

I had a vision yesterday of what the best friendships look like:

Two little boys want to fly. Each crouches on the mulch in one corner of the playground, tugging as hard as they can on their bootlaces, trying to pull themselves up into the air. They tug until veins bulge in their foreheads, but their little boots remain firmly planted on the ground.

One of the boys notices the other, and walks over. After a moment of silence, they each drop their own laces, interweave their arms and hold on to the other's bootlaces. Pulling as hard as they can, they ascend into the air. Faster and faster the boys fly upwards. Until the neon yellow tube slide is the size of a pinky. Until the red brick schoolhouse is the size of an ant. Until the Earth is the size of a droplet of water.

Learn to provide good incentives for the people around you. If the smallest push on a regular basis might solve your problems, providing this push for other people can

solve theirs. And the smallest push in the wrong direction can corrupt the purest of souls. Take a good hard look at the way you interact with people, and what this implies about what you want for them. Are there particular people around whom you happen to always play Devil's Advocate? Are there ways you act to intentionally deceive, manipulate, or ignore?

Laugh at good jokes. Learn when to listen and hold space for others. Give specific praise and gratitude. Provide criticism in a consequentialist manner.

Daily Challenge

Praise me for one thing I've done well in Hammertime and criticize me for one thing I've done badly.

CoZE 3: Empiricism

This is part 25 of 30 in the Hammertime Sequence. Click here for the intro.

> *The boy on the right has gone places. The boy on the left has a map. Whom do you marry?*
>
> ~ Whom

Sometimes, I think that most of the value of the CoZE experiment lies not in the expansion of comfort zones but in the experimental attitude it conveys. A good map-maker must constantly check the territory; the trick is to figure out how.

Day 25: Empiricism

The comfort zone is the region in the environment you understand. It contains the locations you frequent, the skills you have mastered, the people you know well. The farther away you get from your comfort zone, the more unknowns you have to prepare for. The boundaries of the comfort zone are designed to protect you from exactly these dangers: the unknown unknowns.

Overly Conservative Lines

The lines of the comfort zone is very conservatively drawn. In the ancestral environment, mistakes were often fatal: failing a hunt, losing a duel. Even nonfatal mistakes were reproductively so: humiliation in the front of your tribe lasts a long time, and you have no place to run. In this environment, it was reasonable to draw the lines of the comfort zone conservatively, since failure was too costly to test.

What does scientific progress look like in this danger-fraught world? Imagine that every time a scientific experiment fails, the experimenter pays with his life. Science would have progressed much more slowly, if at all.

But the world is no longer as dangerous as the ancestral environment. People live longer, are healthier, and are more mobile between communities. Equally, there are larger and richer positive opportunities outside our comfort zones. These are the preconditions for the viability of the scientific method, and the reason we can now use the power of empiricism, in the form of CoZE experiments, to test our boundaries.

Scientific Detachment

A key realization to make is that the comfort zone is part of your map. That is to say, it makes testable predictions about the territory. Your stage fright is making a testable prediction about how awful the experience of public speaking will be, and how much permanent damage you might sustain from a mistake. Your fear of heights is making a testable prediction about how likely it is for you to fall off a tall ledge unsupported.

Once you understand that the emotional aversions that form the boundaries of your comfort zone are built out of *beliefs about reality*, the logical next step is to design

cheap, safe ways to test those beliefs.

I desire to believe what is true.

Usually, you'll find that the lines of your comfort zone are too simplistic and conservative, and there's obvious ways to tiptoe past it without getting in trouble.

Micro-Experiments

One of the core insights I gleaned from Inadequate Equilibria is that modesty, in the form of Status Regulation and Anxious Overconfidence, is one of the biggest fences around your comfort zone. In that post, Eliezer makes the following recommendation that can't be repeated enough:

> Don't assume you can't do something when it's very cheap to try testing your ability to do it.

Don't assume other people will evaluate you lowly when it's cheap to test that belief.

The comfort zone is a set of beliefs about reality. Test those beliefs.

At very least, take five minutes and try to come up with a cheap experiment to test your beliefs. For example, my light novel Murphy's Quest was a cheap way for me to figure out if it's really true, as my System 1 emphatically stated, I'm terrible at writing fiction.

Design cheap experiments to test your fears.

You're afraid your ideas won't be well-received? Make an anonymous account and post the gentlest form of them.

Let me repeat Eliezer's advice again.

> Don't assume you can't do something when it's very cheap to try testing your ability to do it.

Don't assume other people will evaluate you lowly when it's cheap to test that belief.

One more time.

> Don't assume you can't do something when it's very cheap to try testing your ability to do it.

Don't assume other people will evaluate you lowly when it's cheap to test that belief.

Exercise

Pick something you believe you can't do but haven't checked. Set a Yoda Timer and design a cheap experiment to test this belief.

Pick someone you trust who you believe will evaluate you lowly. Test that belief.

Daily Challenge

Share an experience where you radically underestimated or overestimated your own ability.

Silence

This is part 26 of 30 in the Hammertime Sequence. Click here for the intro.

> 满罐子水不响，半罐子水响叮当
>
> The full can is silent, but the half-empty can makes a loud noise.
>
> ~ Chinese proverb.

Take a bottle or soda can and fill it halfway with water. Shake the can – the water will slosh around loudly.

Now, fill the can to the brim and shake it again. It's almost completely silent.

This is an essay about inner silence – calming one's loudest inner voices to allow quieter voices to speak. Usually, the quieter ones have urgent messages, especially given how long they've been neglected.

This post is, in some sense, a followup to Babble.

An Ocean of Voices

It is common sense that the loudest politician is rarely the wisest. That the child who cries the loudest is rarely the one suffers most. That the friend who criticizes most harshly rarely has the best advice. If anything, the volume of a voice negatively correlates with its value.

The Solitaire Principle states that any failure mode of groups of people also applies within the heart of each single human being. A dozen sub-personalities fight over control of your mind, each of their voices clamoring to drown out the others. Perhaps only one or two of them are consistently allowed to speak.

This picture is further complicated by two features. First, voices are quiet for a reason. There are many things your brain is doing that it doesn't want you to know about (see The Elephant in the Brain). These "meta-cognitive blindspots" may be huge issues in your life that you somehow never get to thinking about. Every time you start, you feel unexpectedly sleepy or preoccupied. Your brain sends an army of louder voices to crowd out the tiny note of confusion whispering: *Look at the elephant! Acknowledge the elephant!*

Second, external voices are also competing for airtime in your head, and may easily drown out even your strongest inner voice, e.g. the phenomenon "the music is so loud I can't hear myself think." All sorts of reading, listening, and watching are processes by which we supplant our internal voices with external ones.

This post is about how attractive and dangerous it is to allow external voices drown internal ones out, once and for all.

The Burden of Consciousness

There are a handful of activities that routinely swallow my time like bottomless holes. Playing video games. Watching anime. Reading fiction. Clicking through Reddit. I feel the urge to throw myself into them periodically.

For a long time, I thought these actions were mainly experiential pica: my brain trying to satisfy my needs for signs of progress, self-improvement, drama, and narrative energy. But the other day, I tried taking a nap instead of watching anime, and it satisfied the same urge. That's when I realized what I was really looking for: the *fast-forward button*.

Living consciously and intentionally was too effortful, facing my problems head-on too painful, and what I wanted more than anything was to shut down my own thoughts and fast-forward through life. Read a thousand-page novel, watch a six-season TV show, scroll through a hundred life stories on AskReddit. These were all ways to forfeit my agency and become a medium for someone else's narrative force.

In sum, the executive thread in my brain did everything in its power to shut itself off.

The Will to Nothingness

The book which for me most poignantly describes the burden of consciousness is Marilynne Robinson's Housekeeping (a novel that I almost don't recommend). It's a depressing story in which every character is on the brink of suicide, philosophically and literally.

Here's a moment when the protagonist's sister Lucille is accused of cheating (emphasis mine):

> Lucille was much too indifferent to school ever to be guilty of cheating, and it was only an evil fate that had prompted her to write Simon Bolivar, and the girl in front of her to write Simon Bolivar, when the answer was obviously General Santa Anna. This was the only error either of them made, and so their papers were identical. Lucille was astonished to find that the teacher was so easily convinced of her guilt, so immovably persuaded of it, calling her up in front of the class and demanding that she account for the identical papers. **Lucille writhed under this violation of her anonymity.** At the mere thought of school, her ears turned red.

This moment clarified for me an insight about exactly the kind of nothingness the girls in *Housekeeping* were after. In this kind of nothingness, apathy, conformity, and anonymity are central, while actual suicide is a mere afterthought.

Following Nietzsche (whom I will presumably never understand), we call this urge the will to nothingness. It prays:

Let me not be heard.

Let me not be seen.

Take away my agency.

Drown out my voice.

Fast-forward me through the years.

Let me be one indistinguishable face in a crowd.

Let not the sunrise bring me joy.

Nor sunset sorrow.

Where does the will to nonexistence come from? Part of it is an insecurity that what you have to say is insufficient, that who you are is too broken to contribute. Part of it is bitterness that the world doesn't deserve to hear your voice and see your face. That these two contradictory ideas coexist in a single heart should only surprise you if you've never met a human being.

The Cure to Nihilism is Silence?!

I will not pretend to know how to solve the problem in general, but this is what worked for me. An insightful friend of mine asked me one question which shook me out of the will to nothingness:

"What if every time you wanted to play video games you just introspected instead?"

It had never occurred to me, despite the fact that I love to write, despite the hours I daydream and doodle at every opportunity, that I could make room for these inner voices by silencing the world completely.

For weeks after that day, I took many long walks, muttering gibberish under my breath. I lay in bed and daydreamed. I wrote for hours without stop. In that time I learned that my will to nothingness was unjustified. I learned that my inner voices would never stop having things to say. Later, I also learned that the world deserves everything I can give it, and more.

Look through your life. What do you do to shut off the burden of consciousness? Do you reach for your phone at boring social engagements? Do you drink or smoke? Do you throw yourself into stories that have little artistic merit just to pass the time?

What would happen if every time you wanted to do that, you introspected instead?

Internal Double Crux

This is part 27 of 30 in the Hammertime Sequence. Click here for the intro.

Focusing is a tool for accessing the messages the many sub-personalities in your subconscious are trying to send you. What happens when two or more of these messages are in conflict with each other?

Internal Double Crux (IDC) is CFAR's answer to this problem. Roughly speaking, it's a script for taking turns Focusing on two conflicting inner voices and holding space for them to debate and compromise. A sort of internal couples therapy, if you will.

Day 27: Internal Double Crux

I had a particularly hard time writing this post, so I'll defer to CFAR's script. Then, I'll list a bunch of points I want to emphasize that one would completely miss reading this script.

It's also possible that what I'm doing is not at all the IDC that CFAR has in mind – in that case, I claim that what I'm doing is also useful.

The IDC Algorithm

Here's the complete script for IDC. It's best to get pen and paper and write down each step, as if you are a neutral observer recording a conversation.

1. Find an internal disagreement

- A "should" that's counter to your current default action
- Something you feel you aren't supposed to think or believe (though secretly you do)
- A step toward your goal that feels useless or unpleasant

2. Operationalize the disagreement

- If there are more than two sides, choose two to start with; focus on what feels important
- Choose names that are charitable and describe the beliefs as they feel from the inside, rather than names that are hostile or judgmental (e.g. the "I deserve rest" side, not the "I'm lazy" side)

3. Seek double cruxes

- *Check for urgency*
  - Is one side more impatient or emotionally salient than the other? Does one side *need* to "speak first"?
  - Is one side more vulnerable to dismissal or misinterpretation (i.e. it's the sort of thing you don't allow yourself to think or feel, because it's wrong or stupid or impractical or vague or otherwise outside of your identity)?

- *Seek an understanding of one side*
  - Let whichever side feels more impatient “explain itself” – why does it feel right or important to react in this way?
  - What things does the *other* side not understand about the world, that this side does? Why can’t the other viewpoint be trusted – what’s bad about letting it call the shots?
- *Seek an understanding of the other side*
  - Check for resonance with what the other side just said – did any of it ring true from the second perspective?
  - What things does the first side not understand about the world? Why can’t it be trusted – why would it be bad if only its priorities were taken into account?

4. Resonate

- Continue to ask each side to speak and summarize the perspective of the other, until both models have incorporated the rationales underlying the other’s conclusions
- Imagine the resolution as an if-then statement, and use your inner sim and other checks to see if either side has any unspoken hesitations about the truth and completeness of that statement

Focusing is the Active Ingredient

Where the script says “focus on what feels important,” it means Focusing. By far the most important step in IDC is finding felt senses for each side of the argument and constructing True Names for them via Focusing.

IDC is a particular type of Focusing centered around alternating between two felt senses, trying to articulate their relationship towards each other. Try to act as the neutral moderator between both these senses, and give each time to speak. During the Resonate step, it is likely that you will experience some kind of “felt shift,” or else the locus of disagreement will change. That is to say, you will uncover via IDC a deeper underlying conflict between the two voices. At this point, take the time to refocus on each side and come up with new names.

The first IDC I tried started with two plainly-named sides “I should floss” and “Flossing is a waste of time.” After further focusing and felt shifts, the two sides sound more like “Flossing is a ritual of self-care showing myself I deserve love” and “Flossing is one of infinitely many impositions by which my parents want to curtail my liberty.” The underlying conflict finally emerges!

To me, the point of IDC is to generate a useful set of Focusing prompts. Internal conflict creates felt senses like nothing else!

Seek Fusion, Not Compromise

As you alternate between the two internal voices, make sure to voice some note of charitability towards the other side. This does not mean that you should compromise naively. In general, you should expect the two sides to both have important data to contribute, and one of the end goals is to learn some general rule which contains each side as special cases.

However emotional the conflict feels, follow this principle: conflicting values are usually based on conflicting beliefs about reality. Each side of your internal conflict has a different set of beliefs about reality which influences the way they believe you should act.

For example, if I tried to start an IDC between the two sides of me saying, respectively, “I want to be more extroverted” and “People are dangerous and awful,” progress might be made by allowing each side to list times human beings have been good and evil to me. Fusion might look like “It’s correct to avoid so-and-so situations and types of people where people act particularly antagonistically, while here are a few specific people I don’t interact with who I obviously want to.”

Fifteen Minutes of IDC

Set a Yoda Timer for 15 minutes. Pick as small of an internal conflict as possible and try to IDC it.

Daily Challenge

In IDC as in life, arguments are rarely about what they seem to be about. Doing the dishes is not about the dishes. Flossing is not about dental care.

Most small conflicts are just battles in raging wars between two giant elephants in the brain. Share an example of this phenomenon that you uncovered through IDC (or otherwise).

Reductionism Revisited

This is part 28 of 30 in the Hammertime Sequence. Click here for the intro.

The last three days of Hammertime, I'll wrap up with some scattered thoughts to reinforce important principles.

Today, I'll return to applications of reductionism to instrumental rationality.

Day 28: Reductionism Revisited

Mysterious Answers: A Brief Review

I had a conversation with a friend in which the topic of comedy popped up briefly. I will strawman his argument to make a point:

> Friend: Well, there's no step-by-step training drill to make someone funny. When I imagine a comedy coach, they probably just ask you to tell jokes and rate how funny they are.
>
> Me: If you didn't know math, would you say the same thing about studying math? That there's no step-by-step approach to teaching induction. Instead, a math teacher just has to let the student try to prove things and rate how rigorous each proof is?
>
> Friend: Point taken.

Irreducible and mysterious complexity, as we know, is a property of the map and not of the territory. It's an easy cognitive mistake to make to believe that many skills, especially the ones you're ignorant of, can't be broken down with reductionism and must instead be learned organically and intuitively.

I think this is a symptom of a general cognitive error that can only be cured by rereading Mysterious Answers half a dozen times. It's important enough to highlight again. The error goes like this:

In my subjective experience, my domain of expertise is concrete and gears-like, amenable to reductionism. I have a detailed mental model of how to go about solving a math problem or writing a blog post, step by atomic step. In my subjective experience, skills I don't have are fuzzy, mysterious, and magical. Training them requires intuition, creativity, and spontaneity. From these defects in the map, I then incorrectly deduce that mysteriousness is an actual property of the territory beyond my competence, i.e. outside my comfort zone.

Mysteriousness is in the mind. Go forth with a Zeno's conviction that all of the territory breaks into infinitesimal pieces, each of which you can individually chew.

Build Form by Cleaning Your Room

> One of the most important things to encourage in the early stages of a new skill is the development of good form. Once you have it, trying harder *works*, whereas if you *don't* have it, trying harder just leads to a lot of frustration and discouragement. And of course, if you have bad habits right from the start, they'll only going to get harder and harder to fix as you ingrain them through practice.
>
> ~ CFAR handbook.

One of the features of a reductionist approach towards instrumental rationality is this: hard problems break into small pieces. Small pieces are easy problems. Therefore, you can get better at solving hard problems by training your cognitive strategies on much easier problems.

True mastery starts with practicing cognitive habits to perfection on exceptionally simple tasks.

Counter-intuitive as this principle may seem, we already know it to be true. We know that students can't move on to algebra until they have perfectly memorized their times tables. We know that before you practice writing you need to master typing or handwriting. In fantasy literature, this idea is ever-present: the novice must spend years levitating a pebble or kindling a flame with perfect control before he moves on to more ambitious magiks.

CFAR calls this principle *building form*, in the sense of physical exercise. (I've been told that) in the weight room, correct lifting form leads to better safety and muscle growth. Learning how to place your feet, tighten your glutes, and arch your back correctly are all important fundamentals to get down before you start benching several plates. All of these fundamentals are best trained on much lighter weights than your current maximum.

Jordan Peterson calls this principle *cleaning your room*. Start by solving the problems in your immediate domain of competence like dust bunnies and unwashed clothes (that reminds me … be right back). It you can't the alignment problem of getting yourself to sleep and wake up on time, expect to hurt yourself trying to save the world.

Also, like the novice's exercise of levitating the pebble, building form is not as simple as it appears. A friend of mine had plans to drop out and apply to work on AI at DeepMind. I told him to fix his sleep schedule first. Two months later, after numerous strategic meetings, he's still working on this problem. At least he's finally recognized its difficulty.

Incremental Progress

Reductionism vs. Procrastination

If you have a procrastination problem, here's a simple cognitive shift based on reductionism that helps. It's a variation on the only piece of "classic self-help advice" I ever found useful. Every time you catch yourself delaying a task to a future date, ask yourself the question:

How much of this task am I willing to do right now?

Answer it honestly. Then, do that much.

Maybe instead of getting exercise, all you're willing to do is go outside for a minute. Maybe instead of filing your taxes, all you're willing to do is organize the relevant forms in a folder. Maybe instead of writing that paper, you can at least tolerate typing in the title and section headers.

The discerning reader will notice this script is essentially a TAP to apply a microscopic CoZE experiment at every task aversion. That's exactly right.

Continuous Grading

Despite how disappointing the project was, I really liked Duncan's Dragon Army Retrospective. One of the tangential reasons for this liking is his use of grades instead of a cruder pass/fail system. Grades imply a smooth, continuous success function which is much easier to optimize.

Human beings are not built to make Fails turn into Passes. Human beings are built to make numbers go up [citation needed].

Score yourself continuously, and you'll have an easier time measuring incremental progress and mentally rewarding yourself for it. Score yourself not for whether or not you did something, but for how much you did and how well you did it.

Daily Challenge

Just now I described a microscopic version of CoZE to apply at the five-second level. How many of the other Hammertime Techniques can you build TAPs to apply minified versions of?

The Strategic Level

This is part 29 of 30 in the Hammertime Sequence. Click here for the intro.

I find myself dragging my feet on the last couple days of each Hammertime cycle. From this and several other data points, I think current my writing attention span is around a week, and drafts and outlines sitting for more than a week feel too stale to finish. Had I known this in advance, I would probably have structured Hammertime as six 5-day sprints.

Reinforcement Learning?

What happens when reinforcement learning isn't enough?

You playing a game of Go against sensei. On move twenty-four, sensei invades your three-space extension with devastating precision, cutting a group you thought was safe into two scattered dragons. The left dragon tries to run away, but sensei cuts its escape route off with a delicate leaning attack on your corner enclosure. It dies with abandon.

The right dragon, now facing the massive wall sensei built up by attacking the left group, tries frantically to make life locally. Its second eye is poked out unceremoniously by a well-placed tesuji. Because of your struggle, sensei has fifty points of territory and thickness radiating across the entire board. You resign.

What is a novice supposed to learn from a game like this? If your teacher leaves you to your own devices to review the game, you might easily conclude any of the following, if not a dozen other things:

1. Don't make three space extensions.
2. Never try to run away.
3. Do not respond to leaning moves.
4. Sacrifice early.
5. Study life and death.

Let's say you learn lesson 1, *don't make three space extensions*. The next week's teaching game, you dutifully plod out two spaces from each approach. Sensei's stones are balanced and efficient while yours are over-concentrated and unimaginative. You lose handily by points.

What happens now? Do you return to three-space extensions, frustrated with two-space ones?

Over-correction and Learning Stopsigns

The Strategic Level is a CFAR flash class about learning strategically: updating in such a way that will actually prevent the same failure modes in the future. The kind of learning above is definitely not strategic.

As I see it, there's two common and overlapping kinds of failure modes in learning, where the lessons learned can be worse than nothing.

The first kind is over-correction:

Had an argument: "I should be more understanding."

Had a panic attack: "I should just care less about everything."

Was a White Knight at Dragon Army: "I should just never trust human beings."

Lost a Go game: "I should never make three-space jumps."

Such overly general lessons can be cures worse than the disease. As your simple strategies progressively fail, you need to come up with and try more and more complicated strategies. You can't just continually bounce between two extremes, refusing to stare the complexity of reality in the face.

The second type of failure is similarly unproductive:

I should have just read out that dan-level life and death problem!

I should have just studied chapter 3 instead of chapter 2!

I should have just tried to use the polynomial method on that problem!

I call these thoughts *learning stopsigns*. A common type of learning stopsign is of the form "should have done so and so," where *so and so* is some arbitrary, brilliant, unreasonable choice you would never have made in advance. Just as semantic stopsigns masquerade as answers, learning stopsigns masquerade as lessons learned while not actually providing practical utility for the future.

The learning stopsign simply says: turn back, nothing to see here, painful thoughts past this point. It's usually accompanied by a nonchalant shrug.

Strategic Learning

What does it mean to learn strategically?

Whenever you fail, try to answer the question, "What way of thinking would I have had to employ to have caught this problem ahead of time?" Every lesson learned is a chance to tune your cognitive strategies to prevent as wide a class of similar problems as possible in the future.

At very least, learn to recognize unproductive over-correction and to drive past learning stopsigns. When you encounter a failure and make a snap judgment about what went wrong, ask yourself: is it any less likely I'll fail in the same way again?

Exercise: Set a Yoda Timer and meditate on your most recent mistakes.

Daily Challenge

Share a story of a cure that was worse than the disease.

Hammertime Final Exam

This is part 30 of 30 in the Hammertime Sequence. Click here for the intro.

One of the overarching themes from CFAR, related to The Strategic Level, is that what you learn at CFAR is not a specific technique or set of techniques, but the cognitive strategy that produced those techniques. It follows that if I learned the right lessons from CFAR, then I would be able to produce qualitatively similar – if not as well empirically tested – new principles and approaches to instrumental rationality.

After CFAR, I wanted to design a test to see if I had learned the right lessons. Hammertime was that sort of test for me. Now here's that same test for you.

The Final Exam

I will give three essay prompts and three difficulty levels. Original ideas would be great, but shining a new light on old hammers is also welcome!

Prompts

1. Design a instrumental rationality technique.
2. Introduce a rationality principle or framework.
3. Describe a cognitive defect, bias, or blindspot.

Difficulty Levels

Bronze Mace mode. Write one essay on one of the topics above.

Steel Cudgel of the Lion mode. Write two of three.

Vorpal Dragonscale Sledgehammer of the Whale mode. Write all three. For each essay, give yourself five minutes to brainstorm and five minutes to write.

Here are my answers.

1. Cooperate First

There's an old story about a famous painter of the Realist school who spent a whole year of his training painting still lives of eggs. Each day, he would draw a single egg over and over. He must have produced thousands of sketches and paintings of eggs. His teacher knew exactly how important fundamentals are.

This same motif is deeply embedded in stories all over the world:

Return to fundamentals. Practice your fundamentals.

The iterated prisoner's dilemma is one of the fundamental lessons of rationality. The world is more like a number of iterated prisoner's dilemmas than you'd think. Human beings are more like tit-for-tat players than you'd think. It follows:

Cooperate First!

The first move you make in any interaction with a new acquaintance should be a cooperate, even if you expect them to defect. Perhaps even if you observe them defecting already.

Here's a lesson I learned from meditating on the maxim Cooperate First:

Cooperating First feels like *accepting an unfair game* from the inside. There will be many situations in life where things are framed in a slightly but noticeably unfair way towards you initially. Err on the side of accepting these games anyway!

2. Below the Object Level

One of my main complaints about rationalists (myself included) is our tendency to escalate to the meta-level too often. For example, in any given discussion, arguments over general discussion norms get much more heated and lively than any discussion of the underlying subject matter. We need to spend more time at the object level, touching reality, making experiments, testing our hypotheses.

The move I use to combat the tendency to escalate meta, I call *looking below the object level*.

Looking below the object level is like the move HPMOR_Harry does to achieve partial transfiguration: continually upping the magnification on your mental microscope to actually stare at the detail in reality. Reality is so exorbitantly detailed it's overwhelming to take it all in. Try.

Look at the folds in your clothes, the way light and shadow play off each other. The way threads interweave. Pinch the cloth and watch the creases reorganize under your fingers.

Now reflect on this fact: falling water is attracted to both positive and negative charges.

What.

There's so much going on under what we think of as the object level.

3. Pre-Excuses

Pre-hindsight is a version of Murphyjitsu where you query your mind for what you will learn from an action in hindsight. Pre-excuses are an unproductive cousin that often derail my work.

As a serial procrastinator, I notice a fairly regular pattern of thinking that appears the couple days before I have to meet a professor, and especially before meeting my thesis adviser. My mind is already spinning excuses on overdrive. Here's what my mind sounds like a full day before I have to meet my adviser, when I think about the meeting:

Sorry, this paper took longer than I expected to read.

Sorry, I was busy from other classes, so I didn't do as much paper-writing as I'd planned to.

Sorry, I got sidetracked by this research problem, so I didn't finish the homework.

That's right, I'm having these thoughts about how to apologize for not doing work *even though I still have plenty of time to do the work*. Even worse, I have these pre-excuse thoughts regularly even if I've done the work expected of me – it feels something like cushioning the fall in case it turns out I did it poorly.

And they're usually not even good excuses.

Hammertime Postmortem

Intro.

A bit less than two months ago, I set out to write about instrumental rationality every day for thirty days. In this post, I will quickly evaluate how well I felt I did along each of my four stated objectives. I will simultaneously evaluate all the Hammertime techniques and ideas by their effectiveness to my life.

This period was my deadline to 80/20 instrumental rationality. Thus, I do not plan to blog any more about it for a while. However, I do want to express my strong intent to write a fourth cycle of Hammertime in the early months of 2019, if only to check my long-term progress.

1. Hammertime Report Card

I will grade myself on the four goals I stated in the first Intermission thread:

> My reasons for writing this sequence were, in clear order of importance: (a) to practice writing, (b) to review CFAR techniques for my own benefit, (c) to entertain, and (d) to teach instrumental rationality.

On reflection, these were equally important goals and I only listed them in that relative order because I believed the later ones would be harder to achieve. I will grade everything out of 100, counting up from zero. Only the relative sizes of the numbers mean anything.

Writing Practice: 90/100

This worked out quite well. I produce content about three times faster than I did at the beginning of Hammertime, with perhaps the slightest decrease in quality. Speed I value as much as strength, so this was an amazing improvement. There are things like organization and style I should have played around with more, and a Yoda Timer of copy-editing after each post would have benefited the writing quality greatly.

Personal CFAR Review: 95/100

Through this process I was forced to reflect on, try out, and push the boundaries of almost every single technique in the manual. Other than a handful of techniques that don't click with me at all, this two-month period has been the perfect amount of time to throw at dedicated instrumental rationality practice. The long-term value of the learning I did at CFAR at least tripled because I did this.

Entertainment: 65/100

Hit or miss. Handful of posts that were really fun to write, and still look fun to read. I noticed a number of clear limitations in my writing toolkit that don't seem to be fixable in a day or two (but might be if I actually tried). Despite my best efforts, I'm still not Eliezer or Scott.

What am I missing? I plan to experiment more with dialogue, which I'm awful at writing but seems to make some of Eliezer's and Scott's funniest stuff. Also, detailed and entertaining expositions of science are sorely missing in my writing – this seems like a gold mine as well.

Teach Instrumental Rationality: 50/100

Not sure this sequence is any better as pedagogical material than just the CFAR Handbook, which is a moderately dry reference manual. Perhaps that's good enough. A handful of people seemed to benefit quite a bit, but my sense is that even among the people who read every post, few did any of the exercises or got any mileage out of this sequence over learning what the concept handles are. In the end, I always made decisions in favor of "write what's interesting for me" rather than "write what I think would be most useful to the reader."

Perhaps an interested reader would like to take a couple hours and reassemble the most useful parts of Hammertime into a cleaner subsequence. As a resource on instrumental rationality instruction at most half of the posts in Hammertime are of high value.

Overall: 75/100

Very impressed with myself that I followed through with this project with only minor delays. Everything went approximately as well as could be Outside-View expected.

My main takeaway is to continue throwing myself headlong into medium-term projects without thinking too much about them, and trust my instincts. It's not obvious that more planning or structure would have helped in net – it may even have soured the whole Hammertime project and caused me not to finish at all.

2. Hammers by Power Level

I will go through the core techniques I covered in Hammertime, and grade them each based on effectiveness in my own life.

I'll sort them into three tiers of awesome. Note that the techniques in Hammertime were already pre-selected from a larger pool of techniques based on how good they seemed to me just after CFAR.

S/A Tier

Focusing: 100/100

Doesn't always work, but when it does … life-changing insights. Probably had three or four over the course of Hammertime. Would recommend.

Yoda Timers: 95/100

Timers and deadlines really up my game. I think I've always shied away from using them because "contest math," "speed," and "competitiveness" became low-status after high school, but man am I built for this. Sometimes I think that if grad school was structured as a serious of olympiads except with open problems, I would get a lot more work done.

Design: 90/100

Amazingly underrated technique. Amortizing everything, allowing myself to remove trivial inconveniences, spending time making my physical space better. Substantially improved my baseline quality of life: sleep quality, overall comfort, aesthetics. If I gave up actively using instrumental rationality right now, the effects of the Design choices I made in the last two months would still last for years.

B/C Tier

Bug Hunt: 80/100

Very useful to practice every so often. Ups your noticing game quite a bit for a long time.

CoZE: 80/100

Another solid technique. Gave me the tools to push through many minor unendorsed aversions and try things instinctively. Doesn't work as well by itself on the bigger aversions – in my experience, these require the aid of Focusing and Focusing is the one doing the work.

Silence: 80/100

I feel as if combating the tendrils of nihilism in everyday life is one of the biggest problems to solve. Silence was my first attempt at framing the problem and offering a partial solution. As always, people need to allow themselves to babble more.

TDT for Humans: 75/100

Important principle that finally allowed me to understand the appeal and utility of virtue ethics/deontology. Requires more iteration and work to make it actionable.

Friendship: 75/100

Noticing the value of and setting up long-term iterated conversations with friends was extremely valuable. Experimenting with this also led me into a handful of awkward social situations and unproductive conversations. I've updated towards there existing even fewer people than I thought with whom I can have interesting conversations on a regular basis.

D/F Tier

Murphyjitsu: 65/100

It feels as painful and difficult to practice as reading ability in Go – life is too chaotic. For now, it's only useful on the five-second level: what are the obvious things that will go wrong? Perhaps after I collect more data about common failure modes Murphyjitsu will be more useful. As of now, I feel woefully uncalibrated.

On the plus side, did inspire my longest work of fiction to date.

TAPs: 60/100

Weird and unnatural to practice. Handful of useful things I thought I installed rapidly faded with time. TAPs seem to last about a week for me without some other regular reinforcement mechanism.

Internal Double Crux: 50/100

Too many steps. The only real value seems to be as a method for generating Focusing targets. This is pretty valuable, but still.

Aversion/Goal Factoring: 30/100

Tried a few times, didn't stick. Much weaker than Focusing. Usually, what I need to do is "find out my true main motive and aversion towards the thing," and once that is done the path forward becomes clear.