# Living Luminously

# Sorting Out Sticky Brains

tl;dr: Just because it doesn't seem like we *should* be able to have beliefs we acknowledge to be irrational, doesn't mean we *don't* have them. If this happens to you, here's a tool to help conceptualize and work around that phenomenon.

There's a general feeling that by the time you've acknowledged that some belief you hold is not based on rational evidence, it has already evaporated. The very act of realizing it's not something you should believe makes it go away. If that's your experience, I applaud your well-organized mind! It's serving you well. This is exactly as it should be.

If only we were all so lucky.

Brains are sticky things. They will hang onto comfortable beliefs that don't make sense anymore, view the world through familiar filters that should have been discarded long ago, see significances and patterns and illusions even if they're known by the *rest* of the brain to be irrelevant. Beliefs *should* be formed on the basis of sound evidence. But that's not the only mechanism we have in our skulls to form them. We're equipped to come by them in other ways, too. It's been observed[1] that believing contradictions is only bad because it entails believing falsehoods. If you can't get rid of one belief in a contradiction, and that's the false one, then believing a contradiction is the best you can do, because then at least you have the true belief too.

The mechanism I use to deal with this is to label my beliefs "official" and "unofficial". My official beliefs have a second-order stamp of approval. I believe them, and I believe that I *should* believe them. Meanwhile, the "unofficial" beliefs are those I can't get rid of, or am not motivated to try really hard to get rid of because they aren't problematic enough to be worth the trouble. They might or might not outright contradict an official belief, but regardless, I try not to act on them.

To those of you with well-ordered minds (for such lucky people seem to exist, if we believe some of the self-reports on this very site), this probably sounds outrageous. If I *know they're probably not true...* And I do. But they still make me expect things. They make me surprised when those expectations are flouted. If I'm asked about their subjects when tired, or not prepared for the question, they'll leap out of my mouth before I can stop them, and they won't feel like lies - because they're not. They're beliefs. I just don't like them very much.

I'll supply an example. I have a rather dreadful phobia of guns, and accordingly, I think they should be illegal. The phobia is a terrible reason to believe in the appropriateness of such a ban: said phobia doesn't even stand in for an informative real experience, since I haven't lost a family member to a stray bullet or anything of the kind. I certainly don't assent to the general proposition "anything that scares me should be illegal". I have no other reasons, except for a vague affection for a cluster of political opinions which includes something along those lines, to believe this belief. Neither the fear nor the affection are reasons I endorse for believing things in general, or this in particular. So this is an unofficial belief. Whenever I can, I avoid acting on it. Until I locate some good reasons to believe something about the topic, I officially have no opinion. I avoid putting myself in situations where I might act on the unofficial belief in the same way I might avoid a store with contents for which I have

an unendorsed desire, like a candy shop.  For instance, when I read about political candidates' stances on issues, I avoid whatever section talks about gun control.

Because I know my brain collects junk like this, I try to avoid making up my mind until I do have a pretty good idea of what's going on.  Once I tell myself, "Okay, I've decided", I run the risk of lodging something permanently in my cortex that won't release its stranglehold on my thought process until kingdom come.  I use tools like "temporarily operating under the assumption that" (some proposition) or declaring myself "unqualified to have an opinion about" (some subject).  The longer I hold my opinions in a state of uncertainty, the less chance I wind up with a permanent epistemic parasite that I have to devote cognitive resources to just to keep it from making me do dumb things.  This is partly because it makes the state of uncertainty come to feel like a default, which makes it simpler to slide back to uncertainty again if it seems warranted.  Partly, it's because the longer I wait, the more evidence I've collected by the time I pick a side, so it's less likely that the belief I acquire is one I'll want to excise in the future.

This is all well and good as a prophylactic.  It doesn't help as much with stuff that snuck in when I was but a mere slip of a youth.  For that, I rely on the official/unofficial distinction, and then *toe the official line as best I can* in thought, word, and deed.  I break in uncomfy official beliefs like new shoes.  You can use your brain's love of routine to your advantage.  Act like you only believe the official beliefs, and the unofficial ones will weaken from disuse.  This isn't a betrayal of your "real" beliefs.  The official beliefs are real too!  They're real, and they're *better*.

[1]I read this in Peter van Inwagen's book "Essay on Free Will" but seem to remember that he got it elsewhere.  I'm not certain where my copy has gotten to lately, so can't check.

# Mental Crystallography

*Brains organize things into familiar patterns, which are different for different people. This can make communication tricky, so it's useful to conceptualize these patterns and use them to help translation efforts.*

Crystals are nifty things! The same sort of crystal will reliably organize in the same pattern, and always break the same way under stress.

Brains are also nifty things! The same person's brain will typically view everything through a favorite lens (or two), and will need to work hard to translate input that comes in through another channel or in different terms. When a brain acquires new concepts - even really vital ones - the new idea will result in recognizeably-shaped brain-bits. Different brains, therefore, handle concepts differently, and this can make it hard for us to talk to each other.

This works on a number of levels, although perhaps the most obvious is the divide between styles of thought on the order of "visual thinker", "verbal thinker", etc. People who differ here have to constantly reinterpret everything they say to one another, moving from non-native mode to native mode and back with every bit of data exchanged. People also store and retrieve memories differently, form first-approximation hypotheses and models differently, prioritize sensory input differently, have different levels of introspective luminosity[1], and experience different affect around concepts and propositions. Over time, we accumulate different skills, knowledge, cognitive habits, shortcuts, and mental filing debris. Intuitions differ - appeals to intuition will only convert people who share the premises natively. We have lots in common, but high enough variance that it's impressive how much we do manage to communicate over not only inferential distances, but also fundamentally diverse brain plans. Basically, you can hit two crystals the same way with the same hammer, but they can still break along different cleavage planes.

This phenomenon is a little like man-with-a-hammer syndrome, which is why I chose that extension of my crystal metaphor. But a person's dependence on their mental crystallography, unlike their wanton use of their hammer, rarely seems to diminish with time. (In fact, youth probably confers some *increased* flexibility - it seems that you can probably train children to have different crystalline structures to some degree, but much less so with adults). MWaH is actually partially explained by the brain's crystallographic regularities. A hammer-idea will only be compelling to you if it aligns with the crystals in your head.

Having "useful" mental crystallography - which lets you comprehend, synthesize, and apply ideas in their most accurate, valuable form - is a type of epistemic luck about the things you can best understand. If you're intrinsically oriented towards mathematical explanations, for instance, and this lets you promptly apprehend the truth and falsity of strings of numbers that would leave my head swimming, you're epistemically lucky about math (while I'm rather likely to be led astray if someone takes the time to put together a plausible verbal explanation that may not match up to the numbers). Some brain structures can use more notions than others, although I'm skeptical that any human has a pure generalist crystal pattern that can make great use of every sort of concept interchangeably without *some* native mode to touch base with regularly.

When you're trying to communicate facts, opinions, and concepts - most especially concepts - it is a useful investment of effort to try to categorize both your audience's crystallography *and your own*. With only one of these pieces of information, you can't optimize your message for its recipient, because you need to know what you're translating *from*, not just have a bead on what you are translating *to*. (If you want to translate the word "blesser" into, say, Tagalog, it might be useful to know if "blesser" is English or French.) And even with fairly good information on both origin and destination, you can wind up with a frustrating disconnect; but given that head start on bridging the gap, you can find wherever the two crystals are most likely to touch with less trial and error.

[1]Introspective luminosity (or just "luminosity") is the subject of a sequence I have planned - this is a preparatory post of sorts. In a nutshell, I use it to mean the discernibility of mental states to their haver - if you're luminously happy, clap your hands.

# Generalizing From One Example

**Related to:** The Psychological Unity of Humankind, Instrumental vs. Epistemic: A Bardic Perspective

"*Everyone generalizes from one example. At least, I do.*"

-- Vlad Taltos (*Issola,* Steven Brust)

My old professor, David Berman, liked to talk about what he called the "typical mind fallacy", which he illustrated through the following example:

There was a debate, in the late 1800s, about whether "imagination" was simply a turn of phrase or a real phenomenon. That is, can people actually create images in their minds which they see vividly, or do they simply say "I saw it in my mind" as a metaphor for considering what it looked like?

Upon hearing this, my response was "How the stars was this actually a real debate? Of course we have mental imagery. Anyone who doesn't think we have mental imagery is either such a fanatical Behaviorist that she doubts the evidence of her own senses, or simply insane." Unfortunately, the professor was able to parade a long list of famous people who denied mental imagery, including some leading scientists of the era. And this was all before Behaviorism even existed.

The debate was resolved by Francis Galton, a fascinating man who among other achievements invented eugenics, the "wisdom of crowds", and standard deviation. Galton gave people some very detailed surveys, and found that some people did have mental imagery and others didn't. The ones who did had simply assumed everyone did, and the ones who didn't had simply assumed everyone didn't, to the point of coming up with absurd justifications for why they were lying or misunderstanding the question. There was a wide spectrum of imaging ability, from about five percent of people with perfect eidetic imagery[1] to three percent of people completely unable to form mental images[2].

Dr. Berman dubbed this the Typical Mind Fallacy: the human tendency to believe that one's own mental structure can be generalized to apply to everyone else's.

He kind of took this idea and ran with it. He interpreted certain passages in George Berkeley's biography to mean that Berkeley was an eidetic imager, and that this was why the idea of the universe as sense-perception held such interest to him. He also suggested that experience of consciousness and qualia were as variable as imaging, and that philosophers who deny their existence (Ryle? Dennett? Behaviorists?) were simply people whose mind lacked the ability to easily experience qualia. In general, he believed philosophy of mind was littered with examples of philosophers taking their own mental experiences and building theories on them, and other philosophers with different mental experiences critiquing them and wondering why they disagreed.

The formal typical mind fallacy is about serious matters of mental structure. But I've also run into something similar with something more like the psyche than the mind: a tendency to generalize from our personalities and behaviors.

For example, I'm about as introverted a person as you're ever likely to meet - anyone

more introverted than I am doesn't communicate with anyone. All through elementary and middle school, I suspected that the other children were out to get me. They kept on grabbing me when I was busy with something and trying to drag me off to do some rough activity with them and their friends. When I protested, they counter-protested and told me I really needed to stop whatever I was doing and come join them. I figured they were bullies who were trying to annoy me, and found ways to hide from them and scare them off.

Eventually I realized that it was a double misunderstanding. They figured I must be like them, and the only thing keeping me from playing their fun games was that I was too shy. I figured they must be like me, and that the only reason they would interrupt a person who was obviously busy reading was that they wanted to annoy him.

Likewise: I can't deal with noise. If someone's being loud, I can't sleep, I can't study, I can't concentrate, I can't do anything except bang my head against the wall and hope they stop. I once had a noisy housemate. Whenever I asked her to keep it down, she told me I was being oversensitive and should just mellow out. I can't claim total victory here, because she was very neat and kept yelling at me for leaving things out of place, and I told her she needed to just mellow out and you couldn't even tell that there was dust on that dresser anyway. It didn't occur to me then that neatness to her might be as necessary and uncompromisable as quiet was to me, and that this was an actual feature of how our minds processed information rather than just some weird quirk on her part.

"Just some weird quirk on her part" and "just being oversensitive" are representative of the problem with the typical psyche fallacy, which is that it's invisible. We tend to neglect the role of differently-built minds in disagreements, and attribute the problems to the other side being deliberately perverse or confused. I happen to know that loud noise seriously pains and debilitates me, but when I say this to other people they think I'm just expressing some weird personal preference for quiet. Think about all those poor non-imagers who thought everyone else was just taking a metaphor about seeing mental images *way* too far and refusing to give it up.

And the reason I'm posting this here is because it's rationality that helps us deal with these problems.

There's some evidence that the usual method of interacting with people involves something sorta like emulating them within our own brain. We think about how we would react, adjust for the other person's differences, and then assume the other person would react that way. This method of interaction is very tempting, and it always feels like it ought to work.

But when statistics tell you that the method that would work on you doesn't work on anyone else, then continuing to follow that gut feeling is a Typical Psyche Fallacy. You've got to be a good rationalist, reject your gut feeling, and follow the data.

I only really discovered this in my last job as a school teacher. There's a lot of data on teaching methods that students enjoy and learn from. I had some of these methods...inflicted...on me during my school days, and I had no intention of abusing my own students in the same way. And when I tried the sorts of really creative stuff I would have loved as a student...it fell completely flat. What ended up working? Something pretty close to the teaching methods I'd hated as a kid. Oh. Well. Now I know why people use them so much. And here I'd gone through life thinking my teachers were just inexplicably bad at what they did, never figuring out that I was just

the odd outlier who couldn't be reached by this sort of stuff.

The other reason I'm posting this here is because I think it relates to some of the discussions of seduction that are going on in MBlume's Bardic thread. There are a lot of not-particularly-complimentary things about women that many men tend to believe. Some guys say that women will never have romantic relationships with their actually-decent-people male friends because they prefer alpha-male jerks who treat them poorly. Other guys say women want to be lied to and tricked. I could go on, but I think most of them are covered in that thread anyway.

The response I hear from most of the women I know is that this is complete balderdash and women aren't like that at all. So what's going on?

Well, I'm afraid I kind of trust the seduction people. They've put a lot of work into their "art" and at least according to their self-report are pretty successful. And unhappy romantically frustrated nice guys everywhere can't be completely wrong.

My theory is that the women in this case are committing a Typical Psyche Fallacy. The women I ask about this are not even remotely close to being a representative sample of all women. They're the kind of women whom a shy and somewhat geeky guy knows and talks about psychology with. Likewise, the type of women who publish strong opinions about this on the Internet aren't close to a representative sample. They're well-educated women who have strong opinions about gender issues and post about them on blogs.

And lest I sound chauvinistic, the same is certainly true of men. I hear a lot of bad things said about men (especially with reference to what they want romantically) that I wouldn't dream of applying to myself, my close friends, or to any man I know. But they're so common and so well-supported that I have excellent reason to believe they're true.

This post has gradually been getting less rigorous and less connected to the formal Typical Mind Fallacy. First I changed it to a Typical Psyche Fallacy so I could talk about things that were more psychological and social than mental. And now it's expanding to cover the related fallacy of believing your own social circle is at least a little representative of society at large, which it very rarely is[3].

It was originally titled "The Typical Mind Fallacy", but I'm taking a hint fromt the quote and changing it to "Generalizing From One Example", because that seems to be the link between all of these errors. We only have direct first-person knowledge one one mind, one psyche, and one social circle, and we find it tempting to treat it as typical even in the face of contrary evidence.

This, I think, is especially important for the sort of people who enjoy Less Wrong, who as far as I can tell are with few exceptions the sort of people who are extreme outliers on every psychometric test ever invented.

**Footnotes**

**1.** Eidetic imagery, vaguely related to the idea of a "photographic memory", is the ability to visualize something and have it be exactly as clear, vivid and obvious as actually seeing it. My professor's example (which Michael Howard somehow remembers even though I only mentioned it once a few years ago) is that although

many people can imagine a picture of a tiger, only an eidetic imager would be able to count the number of stripes.

**2.** According to Galton, people incapable of forming images were overrepresented in math and science. I've since heard that this idea has been challenged, but I can't access the study.

**3.** The example that really drove this home to me: what percent of high school students do you think cheat on tests? What percent have shoplifted? Someone did a survey on this recently and found that the answer was nobhg gjb guveqf unir purngrq naq nobhg bar guveq unir fubcyvsgrq (rot13ed so you have to actually take a guess first). This shocked me and everyone I knew, because we didn't cheat or steal during high school and we didn't know anyone who did. I spent an afternoon trying to find some proof that the study was wrong or unrepresentative and coming up with nothing.

# Living Luminously

**The following posts may be useful background material:** Sorting Out Sticky Brains; Mental Crystallography; Generalizing From One Example

I took the word "luminosity" from "Knowledge and its Limits" by Timothy Williamson, although I'm using it in a different sense than he did. (He referred to "being in a position to know" rather than actually knowing, and in his definition, he doesn't quite restrict himself to mental states and events.) The original ordinary-language sense of "luminous" means "emitting light, especially self-generated light; easily comprehended; clear", which should put the titles into context.

Luminosity, as I'll use the term, is self-awareness. A luminous mental state is one that you have and know that you have. It could be an emotion, a belief or alief, a disposition, a quale, a memory - anything that might happen or be stored in your brain. What's going on in your head? What you come up with when you ponder that question - assuming, nontrivially, that you are accurate - is what's luminous to you. Perhaps surprisingly, it's hard for a lot of people to tell. Even if they can identify the occurrence of individual mental events, they have tremendous difficulty modeling their cognition over time, explaining why it unfolds as it does, or observing ways in which it's changed. With sufficient luminosity, you can inspect your own experiences, opinions, and stored thoughts. You can watch them interact, and discern patterns in how they do that. This lets you predict what you'll think - and in turn, what you'll do - in the future under various possible circumstances.

I've made it a project to increase my luminosity as much as possible over the past several years. While I am not (yet) perfectly luminous, I have already realized considerable improvements in such subsidiary skills like managing my mood, hacking into some of the systems that cause akrasia and other non-endorsed behavior, and simply being less confused about why I do and feel the things I do and feel. I have some reason to believe that I am substantially more luminous than average, because I can ask people what seem to me to be perfectly easy questions about what they're thinking and find them unable to answer. Meanwhile, I'm not trusting my mere impression that I'm generally right when I come to conclusions about myself. My models of myself, after I stop tweaking and toying with them and decide they're probably about right, are borne out a majority of the time by my ongoing behavior. Typically, they'll also match what other people conclude about me, at least on some level.

In this sequence, I hope to share some of the techniques for improving luminosity that I've used. I'm optimistic that at least some of them will be useful to at least some people. However, I may be a walking, talking "results not typical". My prior attempts at improving luminosity in others consist of me asking individually-designed questions in real time, and that's gone fairly well; it remains to be seen if I can distill the basic idea into a format that's generally accessible.

I've divided up the sequence into eight posts, not including this one, which serves as introduction and index. (I'll update the titles in the list below with links as each post goes up.)

- You Are Likely To Be Eaten By A Grue. Why do you want to be luminous? What good does it do, and how does it do it?

- Let There Be Light. How do you get your priors when you start to model yourself, when your existing models are probably full of biases?
- The ABC's of Luminosity. The most fundamental step in learning to be luminous is correlating your affect, behavior, and circumstance.
- Lights, Camera, Action! Luminosity won't happen by itself - you need to practice, and watch out for key mental items.
- The Spotlight. Don't keep your introspection interior. Thoughts are slippery. Label and organize whatever you find in your mind.
- Highlights and Shadows. As you uncover and understand new things about yourself, it's useful to endorse and repudiate your sub-components, and then encourage or interrupt them, respectively.
- City of Lights. It's a handy trick to represent yourself as multiple agents when dealing with tensions in yourself.
- Lampshading. When you have models, test them - but rig your experiments!
- **Bonus posts!**
    - Ureshiku Naritai: A story of how I used luminosity to raise my happiness set point.
    - On Enjoying Disagreeable Company: a luminosity-driven model of how to like people on purpose.
    - Seven Shiny Stories: concrete fictional descriptions of luminosity techniques from this sequence in action. (**NOTE**: Several people remark that SSS dramatically improved their understanding of the sequence. It may be indicated to read each Shiny Story concurrently with its associated post. The Shiny Stories each open with links to the relevant segment, and commenter apophenia has cleverly crossposted the stories under the top posts.)

I have already written all of the posts in this sequence, although I may make edits to later ones in response to feedback on earlier ones, and it's not impossible that someone will ask me something that seems to indicate I should write an additional post. I will dole them out at a pace that responds to community feedback.

# You Are Likely To Be Eaten By A Grue

**Previously in sequence/sequence index:** Living Luminously
**Next in sequence:** Let There Be Light

*Luminosity is fun, useful to others, and important in self-improvement. You should learn about it with this sequence.*

Luminosity? Pah! Who needs it?

It's a legitimate question. The typical human gets through life with astonishingly little introspection, much less careful, accurate introspection. Our models of ourselves are sometimes even worse than our models of each other - we have more data, but also more biases loading up our reflection with noise. Most of the time, most people act on their emotions and beliefs directly, without the interposition of self-aware deliberation. And this doesn't usually seem to get anyone maimed or killed - when was the last time a gravestone read "Here Lies Our Dear Taylor, Who Might Be Alive Today With More Internal Clarity About The Nature Of Memory Retrieval"? Nonsense. If Taylor needs to remember something, it'll present itself, or not, and if there's a chronic problem with the latter then Taylor can export memories to the environment. Figuring out how the memories are stored in the first place and tweaking that is not high on the to-do list.

Still, I think it's worth investing considerable time and effort into improving your luminosity. I submit three reasons why this is so.

First, you are a fascinating creature. It's just plain fun and rewarding to delve into your own mind. People in general are among the most complex, intriguing things in the world. You're no less so. You have lived a fair number of observer-moments. Starting with a native architecture that is pretty special all by itself, you've accumulated a complex set of filters by which you interpret your input - remembered past, experienced present, and anticipated future. You like things; you want things; you believe things; you expect things; you feel things. There's a lot of stuff rolled up and tucked into the fissures of your brain. Wouldn't you like to know what it is? Particularly because it's *you*. Many people find themselves to be their favorite topics. Are you an exception? (There's one way to find out...)

Second, an accurate model of yourself can help you help others deal with you in the best possible way. Right now, they're probably using kludgey agglomerations of self-projection, stereotype, and automatically generated guesses that they may not bother to update as they learn more about you. I'm assuming you don't surround yourself with hostile people who would use accurate data about you to hurt and manipulate you, but if you do, certainly be judicious with whatever information your quest for luminosity supplies. As for everyone else, their having a better model of you will avoid a lot of headaches on everyone's parts. I'll present myself as an example: I hate surprises. Knowing this, and being able to tell a complete and credible story about how this works, I can explain to people who might wish to exchange gifts why they should not spring unknown wrapped items on me, and avoid that source of irritation. Most of the people around me choose not to take actions that they know will irritate me; but without a detailed explanation of exactly *how* my preferences are uncommon, they'll all too easily revert to their base model of a generic person.

Third, and most germane to the remaining posts in this sequence: with a better picture of who you are and what your brain is up to, you can find the best low-hanging fruit in terms of hacks to *change* yourself. If you keep going from point A to point Z, but know nothing about the route in between, then the only way you can avoid a disliked Z is to try to come to a screeching halt right before it happens. If you could monitor the process from the start, and determine what pattern your mind follows along the alphabet, you might find that you can easily intervene at G or Q, and never have to deal with Z again. Similarly, if you try to go from alpha to omega but tend not to wind up at omega, how are you ever going to determine where your obstructions lie unless you pay attention to something other than the bare fact of non-omega? There could be some trivial omicron-related problem that you'd fix in a heartbeat if only you knew it was getting in the way. Additionally, your faulty models of yourself are *already changing you* through such miraculous means as cognitive dissonance. Unless you find out how it's doing that, you lose the chance to monitor and control the process.

An analogy: You're waiting to be picked up at the airport. The designated time comes and goes, and you're sitting by the baggage claim with your suitcases at your feet, your eyes on your watch, and a frown on your face. The person was supposed to pick you up *at the airport*, and isn't there! A clear failure has occurred! But if you phone the person and start screaming "The airport, you fool! I'm at the airport! Why aren't you?" then this will tend not to improve things unless the person never left in the first place out of forgetfulness. If they're stuck in traffic, or were sent out of their way by road construction, or have gotten hopelessly lost, or have been identified by the jackbooted thugs that keep watch at the airport parking lot as a terrorist, reiterating that you had this particular goal in mind won't help. And unless you *find out what is keeping them*, you can't help. You have to know where they are to tell them what detours to take to avoid rush hour; you have to know what diversions were introduced to tell them how to rejoin their planned route; you have to know what landmarks they can see to know where they've gone missing to; you have to know whether to go make Bambi eyes at the security guards and plead misunderstanding. Without rather specific, sensitive data about what's gone wrong, you can't make it right.

In the next posts of this sequence, I'm going to illustrate some methods that have helped me learn more about myself and change what I don't like. With luck, they'll assist you on the project that I've just attempted to convince you to want to undertake.

# Let There Be Light

**Sequence index:** Living Luminously
**Previously in sequence:** You Are Likely To Be Eaten By A Grue
**Next in sequence:** The ABC's of Luminosity

*You can start from psych studies, personality tests, and feedback from people you know when you're learning about yourself. Then you can throw out the stuff that sounds off, keep what sounds good, and move on.*

*You may find your understanding of this post significantly improved if you read the first story from Seven Shiny Stories.*

Where do you get your priors, when you start modeling yourself seriously instead of doing it by halfhearted intuition?

Well, one thing's for sure: not with the caliber of introspection you're most likely starting with. If you've spent any time on this site at all, you know people are riddled with biases and mechanisms for self-deception that systematically confound us about who we are. ("I'm splendid and brilliant! The last five hundred times I did non-splendid non-brilliant things were outrageous flukes!") Humans suck at most things, and obeying the edict "Know thyself!" is not a special case.

The outside view has gotten a bit of a bad rap, but I'm going to defend it - as a jumping-off point, anyway - when I fill our luminosity toolbox. There's a major body of literature designed to figure out just what the hell happens inside our skulls: it's called psychology, and they have a rather impressive track record. For instance, learning about heuristics and biases may let you detect them in action in yourself. I can often tell when I'm about to be subject to the bystander effect ("There is someone sitting in the middle of the road. Should I call 911? I mean, she's sitting up and everything and there are non-alarmed people looking at her - but gosh, I probably don't look alarmed either..."), have made some progress in reducing the extent to which I generalize from one example ("How are you not all driven insane by the spatters of oil all over the stove?!"), and am suspicious when I think I might be above average in some way and have no hard data to back it up ("Now I can be confident that I am in fact good at this sort of problem: I answered all of these questions and most people can't, according to someone who has no motivation to lie!"). Now, even if you *are* a standard psych study subject, of course you aren't going to align with *every psychological finding ever.* They don't even align perfectly with each *other*. But - controlling for some huge, obvious factors, like if you have a mental illness - it's a good place to start.

For narrowing things down beyond what's been turned up as typical human reactions to things, you can try personality tests like Myers-Briggs or Big Five. These are not fantastically reliable sources. However, some of them have some ability to track with some parts of reality. Accordingly, saturate with all the test data you can stand. Filter it for what sounds right ("gosh, I guess I *do* tend to be rather bothered by things out of place in my environment, compared to others") and dump the rest ("huh? I'm not open to experience at all! I won't even try escargot!") - these are rough, first-approximation priors, not posteriors you should actually act on, and you can afford a clumsy process this early in the game. While you're at it, give some thought to your intelligence types, categorize your love language[1] - anything that carves up person-space and puts you in a bit of it.

Additionally, if you have honest friends or relatives, you can ask for their help. Note that even honest ones will probably have a rosy picture of you: they can stand to be around you, so they probably aren't paying excruciatingly close attention to your flaws, and may exaggerate the importance of your virtues relative to a neutral observer's hypothetical opinion. They also aren't around you *all the time*, which will constrict the circumstances in which their model is tested and skew it towards whatever influence their own presence has on you. Their outside perspective is, however, still valuable.

(Tips on getting friends/family to provide feedback: I find musing aloud about myself in an obviously tentative manner to be fairly useful at eliciting some domain-specific input. Some of my friends I can ask point-blank, although it helps to ask about specific situations ("Do you think I'm just tired?" "Was I over the line back there?") rather than general traits that feel more judgmental to discuss ("Am I a jerk?" "Do I use people?"). When you communicate in text and keep logs, you can send people pastes of entire conversations (when this is permissible to your original interlocutor) and ask what your consultant thinks of that. If you do not remember some event, or are willing to pretend not to remember the event, then you can get whoever was with you at the time to recount it from their perspective - this process will automatically paint what you did during the event in the light of outside scrutiny.)

If during your prior-hunting something turns up that seems *wrong* to you, whether it's a whole test result or some specific supposed feature of people in a group that seems otherwise generally fitting, that's great! Now you can rule something out. Think: what makes the model wrong? When have you done something that falsified it? ("That one time last week" is more promising than "back in eighty-nine I think it might have been January".) What are the smallest things you could change to make it sit right? ("Change the word "rapid" to "meticulous" and that's me to a tee!") If it helps, take in the information you gather in small chunks. That way you can inspect them one at a time, instead of only holistically accepting or rejecting what a given test tells you.

If something sounds *right* to you, that's also great! Ask: what predictions does this idea let you make about your cognition and behavior? ("Should you happen to meet a tall, dark stranger, you will make rapid assumptions about his character based on his body language.") How could you test them, and refine the model? (Where do the tall, dark strangers hang out?) If you've behaved in ways inconsistent with this model in the past, what exceptions to the rule does that imply and how can you most concisely, Occam-esque-ly summarize them? ("That one tall, dark stranger was wearing a very cool t-shirt which occluded posture data.")

Nota bene: you may be tempted to throw out things because they sound bad ("I can't be a narcissist! That wouldn't be in keeping with the story I tell about myself!"), rather than because they sound wrong, and to keep things because they sound good ("ooh! I'm funny and smart!"), rather than because they sound right. Recite the Litany of Tarski a few times, if that helps: if you have a trait, you desire to believe that you have the trait. If you do not have a trait, you desire to believe that you do not have the trait. May you not become attached to beliefs you may not want. If you have bad features, knowing about them won't make them worse - and might let you fix, work around, or mitigate them. If you lack good features, deluding yourself about them won't make them appear - and might cost you opportunities to develop them for real. If you *can't answer* the questions "when have you done something that falsified this model?" or "list some examples of times when you've behaved in accordance with

this model" - second guess.  Try again.  Think harder.  You are not guaranteed to be right, and being right should be the aim here.

[1]It looks cheesy, but I've found it remarkably useful as a first-pass approximation of how to deal with people when I've gotten them to answer the question.

# The ABC's of Luminosity

**Sequence index:** Living Luminously
**Previously in sequence:** Let There Be Light
**Next in sequence:** Lights, Camera, Action!

*Affect, behavior, and circumstance interact with each other. These interactions constitute informative patterns that you should identify and use in your luminosity project.*

*You may find your understanding of this post significantly improved if you read the second story from Seven Shiny Stories.*

The single most effective thing you can do when seeking luminosity is to learn to correlate your ABC's, collecting data about how three interrelated items interact and appear together or separately.

A stands for "affect". Affect is how you feel and what's on your mind. It can be far more complicated than "enh, I'm fine" or "today I'm sad". You have room for plenty of simultaneous emotions, and different ones can be directed at different things - being on a generally even keel about two different things isn't the same as being nervous about one and cheerful about the other, and neither state is the same as being entirely focused on one subject that thrills you to pieces. If you're nervous about your performance evaluation but tickled pink that you just bought a shiny new consumer good and looking forward to visiting your cousin next week yet irritated that you just stubbed your toe, all while being amused by the funny song on the radio, that's this. For the sake of the alphabet, I'm lumping in less emotionally laden cognition here, too - what thoughts occur to you, what chains of reasoning you follow, what parts of the environment catch your attention.

B stands for "behavior". Behavior here means *what you actually do*. Include as a dramatically lower-weighted category those things that you *fully intended to do*, and *actually moved to do*, but were then prevented from without from doing, or changed your mind about due to *new, unanticipated information*. This is critical. Fleeting designs and intentions cross our minds continually, and if you don't firmly and definitively place your evidential weight on the things that ultimately result in action, you will get subconsciously cherry-picked subsets of those incomplete plan-wisps. This is particularly problematic because weaker intentions will be dissuaded by minor environmental complications at a much higher rate. Don't worry overmuch about "real" plans that this filtering process discards. You're trying to know yourself *in toto*, not yourself at your best time-slices when you valiantly meant to do good thing X and were buffetted by circumstance: if those dismissed real plans represent typical dispositions you have, then they'll have their share of the cohort of actual behavior. Trust the law of averages.

C stands for "circumstance". This is what's going on around you (what time is it? what's going on in your life now and recently and in the near future - major events, minor upheavals, plans for later, what people say to you? where are you: is it warm, cold, bright, dim, windy, calm, quiet, noisy, aromatic, odorless, featureless, busy, colorful, drab, natural, artificial, pretty, ugly, spacious, cozy, damp, dry, deserted, crowded, formal, informal, familiar, new, cluttered, or tidy?). It also covers what you're doing and things inside you that are generally conceptualized as merely

physical (are you exhausted, jetlagged, drugged, thirsty, hungry, sore, ill, drunk, energetic, itchy, limber, wired, shivering?  are you draped over a recliner, hiding in a cellar, hangliding or dancing or hiking or drumming or hoeing or diving?) Circumstances are a bit easier to observe than affect and behavior.  If you have trouble telling where you are and what you're up to, your first priority shouldn't be luminosity.  And while we often have *some* trouble distinguishing between various physical ailments, there are strong pressures on our species to be able to tell when we're hungry or in pain.  Don't neglect circumstance when performing correlative exercises just because it doesn't seem as "the contents of your skull"-y.  SAD should be evidence enough that our environments *can* profoundly influence our feelings.  And wouldn't it be *weird*, after all, if you felt and acted just the same while ballroom dancing, and while setting the timer on your microwave oven to reheat soup, and while crouching on the floor after having been taken hostage at the bank?

All of these things are interdependent:

- A -> B: Your affect influences your behavior most directly - affect, after all, captures what you're thinking and feeling, and apart from purely reflexive actions, you're going to act in response to that.
- C -> B: Circumstance also feeds very obviously into behavior.  You cannot step on a gas pedal if there isn't one in front of you; you can't take a free sample of tapenade on a cracker if there aren't any to be had; and I've found it dreadfully tricky to start twirling my skirt around when I'm wearing yoga pants.
- A -> C: Affect can change your circumstances via your behavior, but also by altering what happens to your body's condition (we should all be familiar with how stress, for instance, can make you feel physically) and through your fellow human beings by virtue of nonverbal visibility.
- B -> C: Your behavior influences your circumstances, obviously - smash a window, and behold, there is a draft.  Say something and the people around you will probably hear it and react.
- B -> A: Behavior can feed back into affect through hardwired two-way connections (smile and your emotions smile with you!) and through things like consistency effects, which make you become more like the person you seem to behaviorally emulate.
- C -> A: Your circumstances mess with your affect both consciously, through perception and knowledge ("it's my birthday!  yay!"), and subconsciously, via physical effects (if you are operating on excessive sleep debt you will not be pleased with the results).

So don't *just* correlate how they appear together: also note cause and effect relationships.  Until you've developed enough luminosity to detect these things directly, you may have to fall back on a little post-hoc guesswork for connections more complicated than "I was hungry and thinking about cheese, so then I ate some cheese".  Additionally, take note of any interesting *absences*.  If something generally considered sad has happened to you, and you can detect no sadness in your affect or telltale physical side effects, that's highly relevant data.

These correlations will form the building blocks of your first pass of model refinement, proceeding from the priors you extracted from external sources.

# Lights, Camera, Action!

**Sequence index:** Living Luminously
**Previously in sequence:** The ABC's of Luminosity
**Next in sequence:** The Spotlight

*You should pay attention to key mental events, on a regular and frequent basis, because important thoughts can happen very briefly or very occasionally and you need to catch them.*

*You may find your understanding of this post significantly improved if you read the third story from Seven Shiny Stories.*

Luminosity is hard and you are complicated. You can't meditate on yourself for ten minutes over a smoothie and then announce your self-transparency. You have to keep working at it over a long period of time, not least because some effects don't work over the short term. If your affect varies with the seasons, or with major life events, then you'll need to keep up the *first* phase of work through a full year or a major life event, and it turns out those don't happen every alternate Thursday. Additionally, you can't cobble together the best quality models from snippets of introspection that are each five seconds long; extended strings of cognition are important, too, and can take quite a long time to unravel fully.

Sadly, looking at what you are thinking inevitably changes it. With enough introspection, this wouldn't influence your accuracy about your overall self - there's no reason in principle why you couldn't spend all your waking hours noting your own thoughts and forming meta-thoughts in real time - but practically speaking that's not going to happen. Therefore, some of your data will have to come from memory. To minimize the error introduction that comes of retrieving things from storage, it's best to arrange to reflect on very recent thoughts. It may be worth your while to set up an external reminder system to periodically prompt you to look inward, both in the moment and retrospectively over the last brief segment of time. This can be a specifically purposed system (i.e. set a timer to go off every half hour or so), or you can tie it to convenient promptings from the world as-is, like being asked "What's up?" or "Penny for your thoughts".

When you introspect, there is a lot to keep track of. For instance, consider the following:

- What were you thinking about? (This could be more than one thing. You are a massively parallel system.) Was it a concept, image, sensation, desire, belief, person, object, word, place, emotion, plan, memory...?
- How tightly were you focused on it? (Is the topic itself narrow or disparate?) What other items (sensory, cognitive, emotional) seemed to intrude on your concentration, if any, and how did you react to this incursion?
- How did you feel about the subject of the thought? This includes not only emotional reactions like "this is depressing" or "yay!", but also what you felt inclined to do about the topic (if anything), and how important or interesting your thought seemed.
- How does thinking, in general, feel to you? (I conducted an informal survey of this and got no two answers the same. Anecdotally, it may be rather key to determining how you are different from others, and so in refining your model of

yourself relative to the fairly generic priors we're starting with.)  Coming up with a good way to conceptualize your style of thinking can help you interpret introspective data, although be sure to abandon a metaphor that looks about to snap.  You might have different answers when you're "actively" thinking something through - i.e. when novel information is generated in your mind - and when you're thinking "passively", as when you read or listen to some information and absorb its content as it comes.

- What memories did the thought dredge up, if any - parallel situations from the past, apparently unrelated anecdotes that floated by for no reason, events where you learned concepts key to the topic of your thought?  Did the thought generate anticipations for the future - a plan, a fear, a hope, an expectation, a worry?
- What sensory input were you receiving at the time?  Include not only sight, sound, smell, touch, and taste, but also things like temperature, proprioception, and internal symptoms like hunger or nausea.  Can you determine how, if at all, that interacted with the thought?

You cannot have too much data.  (You probably can have too much data in one situation relative to how much you have in another, though - that'll overbalance your models - so make a concerted effort to diversify your times and situations for introspection.)  When you acquire the data, correlate it to learn more about what might bring various aspects of your thought into being.

# The Spotlight

**Sequence index:** Living Luminously
**Previously in sequence:** Lights, Camera, Action
**Next in sequence:** Highlights and Shadows

*Inspecting thoughts is easier and more accurate if they aren't in your head. Look at them in another form from the outside, like they belonged to someone else.*

*You may find your understanding of this post significantly improved if you read the fourth story from Seven Shiny Stories.*

One problem with introspection is that the conclusions you draw about your thoughts are themselves thoughts. Thoughts, of course, can change or disappear before you can extract information about yourself from them. If a flash of unreasonable anger crosses my mind, this might stick around long enough to make me lash out, but then vanish before I discover how unreasonable it was. If thoughts weren't slippery like this, luminosity wouldn't be much of a project. So of course, if you're serious about luminosity, you need a way to pin down your thoughts into a concrete format that will hold still.

You have to pry your thoughts out of your brain.

Writing is the obvious way to do this - for me, anyway. You don't have to publicize what you extract, so it doesn't have to be aesthetic or skillful, just serviceable for your own reference. The key is to get it down in a form that you can look at without having to continue to introspect. Whether this means sketching or scribing or singing, dump your brain out into the environment and have a peek. It's easy to fool yourself into thinking that a given idea makes sense; it's harder to fool someone else. Writing down an idea automatically engages the mechanisms we use to communicate to others, helping you hold your self-analysis to a higher standard.

To turn your thoughts into non-thoughts, use labels to represent them. Put them in reference classes, so that you can notice when the same quale, habit of inference, or thread of cognition repeats. That way, you can detect patterns: "Hey, the last time I felt like this, I said something I really regretted; I'd better watch it." If you can tell when something has happened twice, you can tell when it hasn't - and *new* moods or dispositions are potentially very important. They mean that you or something around you has changed, and that could be a valuable resource or a tricky hazard.

Your labels can map onto traditional terms or not - if you want to call the feeling of having just dropped your ice cream on the sidewalk "blortrath", no one will stop you. (It can be useful, later when you're trying to share your conclusions about yourself with others, to have a vocabulary of emotion that overlaps significantly with theirs; but you can always set up an idiolect-to-dialect dictionary later.) I do recommend identifying labeled items as being more or less similar to each other (e.g. annoyance is more like fury than it is like glee) and having a way to account for that in your symbolism. Similarities like that will make it more obvious how you can generalize strategies from one thing to another.

Especially if you don't think in words, you might find it challenging to turn your thoughts into something in the world that represents them. Maybe, for instance, you

think in pictures but aren't at all good at drawing.  This is one of the steps in luminosity that I think is potentially dispensible, so if you honestly cannot think of any way to jot down the dance of your mind for later inspection, you can just work on thinking very carefully such that if something were to be out of place the next time you came back to your thought, you'd notice it.  I do recommend spending at least five to ten minutes trying to write, diagram, draw, mutter, or interpretive-dance your mental activity before you give it up as untenable for you, however.

Once you have produced a visible or audible translation of your thoughts, *analyze it the way you would if someone else had written it*.  (Except inasmuch as it's in a code that's uniquely understandable to you and you shouldn't pretend to do cryptanalysis on it.)  What would you think of the person described if you didn't know anything else?  How would you explain these thoughts?  What threads of reasoning seem to run in the background from one belief to another, or from a perception to a belief, or from a desire to an intention?  What do you expect this person to do next?  What's your next best guess after that?  And: what more do you want to know?  If you met the person described, how could you satisfy your curiosity without relying on the bias-laden answer you'd get in response to a verbal inquiry?  Try it now - in a comment under this post, if you like: note what you're thinking, as much of it as you can grab and get down.  Turn on the anti-kibitzer and pretend someone else said it: what must be going on in the mind behind this writing?

# Highlights and Shadows

**Sequence index:** Living Luminously
**Previously in sequence:** The Spotlight
**Next in sequence:** City of Lights

*Part of a good luminosity endeavor is to decide what parts of yourself you do and don't like.*

*You may find your understanding of this post significantly improved if you read the fifth story from Seven Shiny Stories.*

As you uncover and understand new things about yourself, you might find that you like some of them, but don't like others. While one would hope that you'd be generally pleased with yourself, it's a rare arrogance or a rarer saintliness that would enable unlimited approval. Fortunately, as promised in post two, luminosity can let you determine what you'd like to change as well as what's already present.

But what to change?

An important step in the luminosity project is to sort your thoughts and feelings not only by type, correlation, strength, etc, but also by *endorsement*. You endorse those thoughts that you like, find representative of your favorite traits, prefer to see carried into action, and wish to keep intact (at least for the duration of their useful lives). By contrast, you repudiate those thoughts that you dislike, consider indicative of negative characteristics, want to keep inefficacious, and desire to modify or be rid of entirely.

Deciding which is which might not be trivial. You might need to sift through several orders of desire before finally figuring out whether you want to want cake, or like liking sleep, or prefer your preference for preferentism. A good place to start is with your macro-level goals and theoretical commitments (e.g., when this preference is efficacious, does it serve your Life Purpose™, directly or indirectly? if you have firm metaethical notions of right and wrong, is this tendency you have uncovered in yourself one that impels you to do right things?).

As a second pass, you can work with the information you collected when you correlated your ABCs. How does an evaluated desire makes you feel when satisfied or unsatisfied? Does it cripple you when unsatisfied or improve your performance when satisfied? Are you reliably in a position to satisfy it? If you can't typically satisfy it, would it be easier to change the desire or to change the circumstances that prevent its satisfaction? However, this is a second step. You need to know what affect and behavior are preferable to you before you can judge desires (and other mental activity) relative to what they yield in those departments, and judging affect and behavior is itself an exercise in endorsement and repudiation.

Knowing what you like and don't like about your mind is a fine thing. Once you have that information, you can put it to direct use immediately - I find it useful to tag many of my expressions of emotion with the words "endorsed" or "non-endorsed". That way, the people around me can use that categorization rather than having to either assume I approve of everything I feel, or layer their own projections of endorsement on top of me. Either would be unreliable and cause people to have poor models of me: I have not yet managed to excise my every unwanted trait, and my patterns of

endorsement do not typically map on to the ones that the people around me have or expect me to have.

Additionally, once you know what you like and don't like about your mind, you can begin to make progress in increasing the ratio of liked to unliked characteristics. People often make haphazard lurches towards trying to be "better people", but when "better" means "lines up more closely with vaguely defined commonsense intuitions about morality", this is not the sort of goal we're at all good at pursuing. Specific projects like being generous or more mindful are a step closer, but the greatest marginal benefit in self-revision comes of figuring out what comes in *advance* of behaving in a non-endorsed way and heading it off at the pass. (More on this in "Lampshading".) The odds are low that your brain's patterns align closely with conventional virtues well enough for them to be useful targets. It's a better plan to identify what's already present, then endorse or repudiate these pre-sliced thoughts and work on them as they appear instead of sweeping together an unnatural category.

# City of Lights

**Sequence index:** Living Luminously
**Previously in sequence:** Highlights and Shadows
**Next in Sequence:** Lampshading

*Pretending to be multiple agents is a useful way to represent your psychology and uncover hidden complexities.*

*You may find your understanding of this post significantly improved if you read the sixth story from Seven Shiny Stories.*

When grappling with the complex web of traits and patterns that is you, you are reasonably likely to find yourself less than completely uniform. You might have several competing perspectives, possess the ability to code-switch between different styles of thought, or even believe outright contradictions. It's bound to make it harder to think about yourself when you find this kind of convolution.

Unfortunately, we don't have the vocabulary or even the mental architecture to easily think of or describe ourselves (nor other people) as containing such multitudes. The closest we come in typical conversation more resembles descriptions of superficial, vague ambivalence ("I'm sorta happy about it, but kind of sad at the same time! Weird!") than the sort of deep-level muddle and conflict that can occupy a brain. The models of the human psyche that have come closest to approximating this mess are what I call "multi-agent models". (Note: I have no idea how what I am about to describe interacts with actual psychiatric conditions involving multiple personalities, voices in one's head, or other potentially similar-sounding phenomena. I describe multi-agent models as employed by psychiatrically singular persons.)

Multi-agent models have been around for a long time: in Plato's *Republic*, he talks about appetite (itself imperfectly self-consistent), spirit, and reason, forming a tripartite soul. He discusses their functions as though each has its own agency and could perceive, desire, plan, and act given the chance (plus the possibility of one forcing down the other two to rule the soul unopposed). Not too far off in structure is the Freudian id/superego/ego model. The notion of the multi-agent self even appears in fiction (warning: TV Tropes). It appears to be a surprisingly prevalent and natural method for conceptualizing the complicated mind of the average human being. Of course, talking about it as something to *do* rather than as a way to push your psychological theories or your notion of the ideal city structure or a dramatization of a moral conflict makes you sound like an insane person. Bear with me - I have data on the usefulness of the practice from more than one outside source.

There is no reason to limit yourself to traditional multi-agent models endorsed by dead philosophers, psychologists, or cartoonists if you find you break down more naturally along some other arrangement. You can have two of you, or five, or twelve. (More than you can keep track of and differentiate is not a recommended strategy - if you're very tempted to go with this many it may be a sign of something unhealthful going on. If a group of them form a reliable coalition it may be best to fold them back into each other and call them one sub-agent, not several.) Stick with a core ensemble or encourage brief cameos of peripheral aspects. Name them descriptively or after structures of the brain or for the colors of the rainbow, as long as you can tell them apart. Talk to yourselves aloud or in writing, or just think through the interaction if

you think you'll get enough out of it that way.  Some examples of things that could get their own sub-agents include:

- Desires or clusters of desires, be they complex and lofty ("desire for the well being of all living things") or simple and reptilian ("desire for cake")
- "Inner child" or similar role-like groupings of traits ("professional me", "family-oriented me", "hobbies me")
- High-order dispositions and principles ("conscience", "neuroticism", "sense of justice")
- Opinions or viewpoints, either specific to a situation or general trends ("optimism", "outside view", "I should do X")
- Initially unspecified, gradually-personality-developing sub-agents, if no obvious ones present themselves (named for something less suggestive like cardinal directions or two possible nicknames derived from your name)

By priors picked up from descriptions of various people trying this, you're reasonably likely to identify one of your sub-agents as "you".  In fact, one sub-agent may be *solely* identified as "you" - it's very hard to shake the monolithic observer experience. This is fine, especially if the "you" sub-agent is the one that endorses or repudiates, but don't let the endorsement and repudiation get out of hand during multi-agent exercises.  You have to deal with all of your sub-agents, not just the one(s) you like best, and sub-agents have been known to exhibit manipulative and even vengeful behaviors once given voice - i.e. if you represent your desire for cake as a sub-agent, and you have been thwarting your desire for cake for years, you might find that Desire For Cake is pissed off at Self-Restraint and says mean things thereunto.  It will not placate Desire For Cake for you to throw in endorsement behind Self-Restraint while Desire For Cake is just trying to talk to you about your desperate yen for tiramisu. Until and unless you understand Desire For Cake well enough to surgically remove it, you need to work with it.  Opposing it directly and with normative censure will be likely to make it angry and more devious in causing you to eat cake.

A few miscellaneous notes on sub-agents:

Your sub-agents may surprise you far more than you expect to be surprised by... well... yourself, which is part of what makes this exercise so useful.  If you consciously steer the entire dialogue you will not get as much out of it - then you're just writing self-insert fanfiction about the workings of your brain, not actually learning about it.

Not all of your sub-agents will be "interested" in every problem, and therefore won't have much of relevance to say at all times.  (Desire For Cake probably couldn't care less how you act on your date next week until it's time to order dessert.)

Your sub-agents should not outright  lie to each other ("should" in the predictive, not normative, sense - let me know if it turns out yours do), but they may threaten, negotiate, hide, and be genuinely ignorant about themselves.

Your sub-agents may not all communicate effectively.  Having a translation sub-agent handy could be useful, if they are having trouble interpreting each other.

(Post your ensemble of subagencies in the comments, to inspire others!  Write dialogues between them!)

# Lampshading

**Sequence index:** Living Luminously
**Previously in sequence:** City of Lights

*You can use luminosity to help you effectively change yourself into someone you'd more like to be. Accomplish this by fixing your self-tests so they get good results.*

*You may find your understanding of this post significantly improved if you read the seventh story from Seven Shiny Stories.*

When you have coherent models of yourself, it only makes good empirical sense to put them to the test.

Thing is, when you run a test on yourself, you know what test you're running, and what data would support which hypothesis. All that and you're the subject generating the data, too. It's kind of hard to have good scientific controls around this sort of experiment.

Luckily, it turns out that for this purpose they're unnecessary! Remember, you're not just trying to determine what's going on in a static part of yourself. You're also evaluating and changing the things you repudiate when you can. You don't just have the chance to let knowledge of your self-observation nudge your behavior - you can outright *rig your tests*.

Suppose that your model of yourself predicts that you will do something you don't think you should do - for instance, suppose it predicts that you will yell at your cousin the next time she drops by and tracks mud on your carpet, or something, and you think you ought not to yell. Well, you can falsify that model which says you'll yell by not yelling: clearly, if you do not yell at her, then you cannot be accurately described by any model that predicts that you'll yell. By refraining from yelling you push the nearest accurate model towards something like "may yell if not careful to think before speaking" or "used to yell, but has since grown past that". And if you'd rather be accurately described by one of those models than by the "yells" model... you can *not yell*.

(Note, of course, that falsifying the model "yells" by silently picking up your cousin and defenestrating her is not an *improvement*. You want to replace the disliked model with a more likable one. If it turns out that you cannot do that - if controlling your scream means that you itch so badly to fling your cousin out a window that you're likely to actually do it - then you should postpone your model falsification until a later time.)

Now, of course figuring out how to not yell (let us not forget akrasia, after all) will be easier once you have an understanding of what would make you do it in the first place. Armed with that, you can determine how to control your circumstances to prevent yelling-triggers from manifesting themselves. Or, you can attempt the more difficult but more stable psychic surgery that interrupts the process from circumstance to behavior.

Sadly, I can't be as specific as would be ideal here because so much depends on the exact habits of *your* brain as opposed to any other brains, including mine. You may

need to go through various strategies before you hit on one that works for you to change what you need to change.  You could find that successful strategies eventually "wear off" and need replacing and their edifices rebuilding.  You might find listening to what other people do helpful (post techniques below!) - or you might not.

# Ureshiku Naritai

This is a supplement to the luminosity sequence. In this comment, I mentioned that I have raised my happiness set point (among other things), and this declaration was met with some interest. Some of the details are lost to memory, but below, I reconstruct for your analysis what I can of the process. It contains *lots of gooey self-disclosure*; skip if that's not your thing.

In summary: I decided that I had to and wanted to become happier; I re-labeled my moods and approached their management accordingly; and I consistently treated my mood maintenance and its support behaviors (including discovering new techniques) as immensely important. The steps in more detail:

1. *I came to understand the **necessity** of becoming happier.* Being unhappy was not just unpleasant. It was *dangerous*: I had a history of suicidal ideation. This hadn't resulted in actual attempts at killing myself, largely because I attached hopes for improvement to concrete external milestones (various academic progressions) and therefore imagined myself a magical healing when I got the next diploma (the next one, the next one.) Once I noticed I was doing that, it was unsustainable. If I wanted to live, I had to find a safe emotional place on which to stand. It had to be my top priority. This required several sub-projects:

- I had to eliminate the baggage that told me it was *appropriate* or *accurate* to feel bad most of the time. I endorse my ability to react emotionally to my environment; but this should be acute, not chronic. Reacting emotionally is about feeling worse when things *get* worse, not feeling bad when things are bad for months or years on end. (Especially not when feeling bad reduces the ability to make things less bad.) Further, having a lower set point did not affect my emotional *range* except to shrink it; it reduced the possible impact of real grief, and wasn't compatible with the "react emotionally" plan. The low set point also compromised my ability to react emotionally to *positive* input, because it was attached to a systematic discounting of such positivity.
- I had to eliminate the baggage that told me it was not possible to *cognitively* change my mood. Moods correspond to thoughts, and while it can be hard to *avoid* thinking about things, I can decide *to* think about whatever I want. A decade of assorted antidepressants had wreaked no discernible change on my affect, which constituted strong evidence that chemicals were not my problem. And it was easy to see that my mood varied on a small scale with things under my complete or partial control, like sleep, diet, and activity. It did not seem outrageous that long-term, large-scale interventions could have similar effects on my overall mood.
- I had to decide, and act on the decision, that my happiness was important and worth my time and attention. I had to pay attention, and note what helped and what hurt. I had to put increasing the helping factors and decreasing the hurting factors at the top of my list whenever it was remotely feasible, and relax my standards around "remote feasibility" to prevent self-sabotage. And I had to commit to abandoning counterproductive projects or interactions, at least until I'd developed the stability to deal with the emotions they generated without suffering permanent setbacks.

2. *I **re-labeled** my moods, so that identifying them in the moment prompted the right actions.* When a given point on the unhappy-happy spectrum - let's call it "2" on

a scale of 1 to 10 - was labeled "normal" or "set point", then when I was feeling "2", I didn't assume that meant anything; that was the default state. That left me feeling "2" a lot of the time, and when things went wrong, I dipped lower, and I waited for things outside of myself to go right before I went higher. The problem was that "2" was not a good place to be spending most of my time.

- I had to label the old set-point as subnormal, a *problem state* that generated a need for immediate action from me to *fix* it. It was like telling myself that, unbeknownst to me, my left foot was in constant pain and needed medicine at once: kind of hard to swallow, given that my left foot always felt pretty much the same unless I'd just stubbed a toe or received a massage. But eventually, I attached *urgency* to the old set point. It was not *just* how things were normally; it was a sign that something was *wrong.*
- I had to make sure that I had many accessible, cheap excuses to cheer up, so I didn't ever fall into the trap of "just this once" leaving myself at a "2" state instead of acting. I designated a favorite pair of socks and wore them whenever I woke up on the wrong side of the bed; I took up the habit of saving every picture of a cute animal I found on the Internet so I could leaf through the collection whenever I wanted; I threw myself into developing the skill of making friends *on purpose* so I'd have lots and if I happened to log onto my IM client, someone would be there who would talk to me; I became very acquisitive of inexpensive goods like music and interesting websites. When one of these interventions failed to work, I forced myself to try something else, rather than falling into the self-talk disaster of "well, that didn't help; I guess something must really be wrong and I should feel like this until it goes away by itself." I also harnessed my tendency to feel better after a night's sleep - if I felt suboptimal close to bedtime, I'd turn in early and reasonably expect to wake up improved.
- I stopped tolerating the minor injuries to my affect that I identified as most consistent and, therefore, most likely to contribute to my poor set point. For instance, I noticed that I always slept better when I didn't go to bed expecting to awaken to the sound of an alarm, so I aggressively rearranged my schedule to give me morning leeway, and found alarm software that would wake me more gently when an early start was absolutely necessary. I identified people with whom interaction was frustrating and draining, and I limited interaction with them both by reducing opportunities to start, and by dropping my standards for abandoning the exchange midway through so I could leave before things got very bad. I practiced, in general, "writing things off" and rehearsed internal monologues about how I no longer needed to worry about [thing X]. ("I cannot control the speed of the bus. I caught it, and it will get there when it gets there. There is no point in further fretting about being late until I'm moving under my own power again - so I'll stop. To manage my strong, intrusive desire to be on time, I will start thinking about how to choose an efficient path to walk once I get off the bus.")
- I labeled my new desired set point - a safe spot on the spectrum, call it "5", which was ambitious yet felt attainable - as "normal". When asked how I was in this state, I consciously chose to say that I was "fine" or "okay" instead of something more enthusiastic, like "great", that I might have said before - the energy I felt at "5" was no longer to be considered *extra*. Similarly, these were not suitable occasions to do displeasing things. I didn't have happiness to burn at "5" - I waited until I was even better before I relaxed my emotional avarice. Instead, "5" was a good place from which to undertake more expensive entertainments that offered net improvement. (More difficult than choosing a specific pair of socks to wear is starting a D&D game, or walking around and

exploring a new location, or working on a piece of artwork or fiction; the lag time and effort makes them poor "cheer up" activities, but excellent ways to get from "5" to "6" or "7".)
- I made a point of noting non-sadness deficiencies in my status like boredom, hunger, tiredness, or annoyance. These weren't directly related to the set point I was trying to affect, but they could exacerbate a bad influence or limit the power of a good one. Additionally, at the level of luminosity I then had to work with, they could also mask moods that were actually sadness, in much the same way that sometimes one can feel hungry when in fact just thirsty.

3. *I treated my own mood as **manageable***. Thinking of it as a thing that attacked me with no rhyme or reason - treating a bout of depression like a cold - didn't just cost me the opportunity to fight it, but also made the entire situation seem more out-of-control and hopeless. I was wary of learned helplessness; I decided that it would be best to interpret my historically static set point as an indication that I hadn't hit on the right techniques yet, not as an indication that it was inviolable and everlasting. Additionally, the fact that I didn't know how to fix it yet meant that if it was going to be my top priority, I had to treat the value of information as very high; it was worth experimenting, and I didn't have to wait for surety before I gave something a shot.

- Even if I determined that my mood reacted to my environment in some way, that only removed my power over it one step: I could control my environment to a considerable degree, and with a strong enough reason to do so, I committed to enacting that power. (This sometimes has had unexpected and dramatic consequences. For example, once I determined that grad school was no longer compatible with my happiness, I dropped out as soon as I had something promising to switch to - mid semester - and moved across the country. To excellent effect, I might add.)
- Even if I have a lot on my plate, being happier will help me do it. It's like sleep: it's easy to keep staying up and staying up, because sleep just seems so *unproductive*, and you can get *some* work done however tired you are. But over the long term, getting to sleep at a sane hour every day will let you accomplish more; and so with maintaining a good affect consistently. Mood maintenance is typically not the most *immediately productive* thing I could be doing, but treating it as my top priority save in dire emergency has let me be more effective than I was before.
- I had to be willing to expend resources on my project. This involved working around some neuroses, like my unwillingness to spend money, and overcoming some background reluctance to try new things. Also, I had to allow myself to be somewhat subject to my whims. I still don't know what makes the mood to, say, do artwork strike me, but when it strikes, I have to do art or lose the inclination. Efficacious inclinations to do fun things are precious to me, and so whenever possible, I don't restrain them - even though this costs time and occludes other activities.

# A Suite of Pragmatic Considerations in Favor of Niceness

tl;dr: Sometimes, people don't try as hard as they could to be nice. If being nice is not a terminal value for you, here are some other things to think about which might induce you to be nice anyway.

There is a prevailing ethos in communities similar to ours - atheistic, intellectual groupings, who congregate around a topic rather than simply to congregate - and this ethos says that it is not necessary to be nice. I'm drawing on a commonsense notion of "niceness" here, which I hope won't confuse anyone (another feature of communities like this is that it's very easy to find people who claim to be confused by monosyllables). I do not merely mean "polite", which can be superficially like niceness when the person to whom the politeness is directed is in earshot but tends to be far more superficial. I claim that this ethos is mistaken and harmful. In so claiming, I do not also claim that I am always perfectly nice; I claim merely that I and others have good reasons to try to be.

The dispensing with niceness probably springs in large part from an extreme rejection of the ad hominem fallacy and of emotionally-based reasoning. Of course someone may be entirely miserable company and still have brilliant, cogent ideas; to reject communication with someone who just happens to be miserable company, in spite of their brilliant, cogent ideas, is to miss out on the (valuable) latter because of a silly emotional reaction to the (irrelevant) former. Since the point of the community is ideas; and the person's ideas are good; and how much fun they are to be around is irrelevant - well, bringing up that they are just terribly mean seems trivial at best, and perhaps an invocation of the aforementioned fallacy. We are here to talk about *ideas*! (Interestingly, this same courtesy is rarely extended to appalling spelling.)

The ad hominem fallacy *is* a fallacy, so this is a useful norm up to a point, but not up to the point where people who are perfectly capable of being nice, or learning to be nice, neglect to do so because it's apparently been rendered locally worthless. I submit that there are still good, pragmatic reasons to be nice, as follows. (These are claims about how to behave around real human-type persons. Many of them would likely be obsolete if we were all perfect Bayesians.)

1. It provides good incentives for others. It's easy enough to develop purely subconscious aversions to things that are unpleasant. If you are miserable company, people may stop talking to you without even knowing they're doing it, and some of these people may have ideas that would have benefited you.
2. It helps you hold off on proposing diagnoses. As tempting as it may be to dismiss people as crazy or stupid, this is a dangerous label for us biased creatures. Fewer people than you are tempted to call these things are genuinely worth writing off as thoroughly as this kind of name-calling may tempt you to do. Conveniently, both these words (as applied to people, more than ideas) and closely related ones are culturally considered mean, and a general niceness policy will exclude them.
3. It lets you exist in a cognitively diverse environment. Meanness is more tempting as an earlier resort when there's some kind of miscommunication, and miscommunication is more likely when you and your interlocutor think differently. Per #1, not making a conscious effort to be nice will tend to drive off

the people with the greatest ratio of interesting new contributions to old rehashed repetitions.

4. It is a cooperative behavior. It's obvious that it's nicer to live in a world where everybody is nice than in a world where everyone is a jerk. What's less obvious, but still, I think, true, is that the cost of cooperatively being nice while others are mean is in fact *very low*. This is partly because human interaction is virtually always iterated, (semi-)public, or both; and also because it's just not very hard to be nice. The former lets you reap an excellent signaling effect:
5. It signals the hell out of your maturity, humility, and general awesome. If you spend as much time on the Internet as I do, you read a few online content publishers who publicly respond to their hate mail. It can sometimes be funny to read the nasty replies. But I generally walk away thinking more of the magnanimous ones who are patient even with their attackers.
6. It promotes productive affect in yourself and others. The atmosphere of a relationship or group has many effects, plenty of which aren't cognitively luminous, and some of which can spill over into your general mood and whatever you were hoping to use your brain for.
7. It is useful in theoretical discussions to draw a distinction between being mean to someone and doing something that's seriously morally wrong, but this line is fuzzier or completely absent in human prephilosophical intuitions. If you are ever troubled by ethical akrasia, it may be easier to stave off if you try to avoid delivering small slights and injuries as well as large violations.
8. It yields resources in the form of friendly others. Whether you are an introvert or an extrovert, other people can be useful to have around, and not even just for companionship. Compared to indifferent or actively hostile neighbors, it's an obvious win to be nice and win what goodwill you can.
9. It can save time - often yielding a net benefit, rather than wasting time as is sometimes complained. For instance, if a miscommunication is made, a mean response is to interpret the misstatement at face value and ridicule or attack - this can devolve into a time-consuming fight and may never resolve the initial issue. A nice response is to gently clarify, which can be over in minutes.

# On Enjoying Disagreeable Company

**Bears resemblance to:** Ureshiku Naritai; A Suite of Pragmatic Considerations In Favor of Niceness

*In this comment, I mentioned that I can like people on purpose. At the behest of the recipients of my presentation on how to do so, I've written up in post form my tips on the subject. I have not included, and will not include, any specific real-life examples (everything below is made up), because I am concerned that people who I like on purpose will be upset to find that this is the case, in spite of the fact that the liking (once generated) is entirely sincere. If anyone would find more concreteness helpful, I'm willing to come up with brief fictional stories to cover this gap.*

It is useful to like people. For one thing, if you have to be around them, liking them makes this far more pleasant. For another, well, they can often *tell*, and if they know you to like them this will often be instrumentally useful to you. As such, it's very handy to be able to like someone you want to like deliberately when it doesn't happen by itself. There are three basic components to liking someone on purpose. First, reduce salience of the disliked traits by separating, recasting, and downplaying them; second, increase salience of positive traits by identifying, investigating, and admiring them; and third, behave in such a way as to reap consistency effects.

**1. Reduce salience of disliked traits.**

Identify the traits you don't like about the person - this might be a handful of irksome habits or a list as long as your arm of deep character flaws, but make sure you know what they are. Notice that however immense a set of characteristics you generate, it's not the entire person. ("Everything!!!!" is not an acceptable entry in this step.) No person can be fully described by a list of things you have noticed about them. Note, accordingly, that you dislike *these things* about the person; but that this does not logically entail disliking *the person*. Put the list in a "box" - *separate* from how you will eventually evaluate the person.

When the person exhibits a characteristic, habit, or tendency you have on your list (or, probably just to aggravate you, turns out to have a new one), be on your guard immediately for the fundamental attribution error. It is especially insidious when you already dislike the person, and so it's important to compensate consciously and directly for its influence. Elevate to conscious thought an "attribution story", in which you consider a *circumstance* - not a character trait - which would explain this most recent example of bad behavior.[1] This should be the most likely story you can come up with that doesn't resort to grumbling about how dreadful the person is - that is, don't resort to "Well, maybe he was brainwashed by Martians, but sheesh, how likely is *that?*" Better would be "I know she was up late last night, and she does look a bit tired," or "Maybe that three-hour phone call he ended just now was about something terribly stressful."

Reach a little farther if you don't have this kind of information - "I'd probably act that way if I were coming down with a cold; I wonder if she's sick?" is an acceptable speculation even absent the least sniffle. If you can, it's also a good idea to *ask* (earnestly, curiously, respectfully, kindly! not accusatively, rudely, intrusively, belligerently!) why the person did whatever they did. Rest assured that if their psyche is fairly normal, an explanation exists in their minds that doesn't boil down to "I'm a

lousy excuse for a person who intrinsically does evil things just because it is my nature." (Note, however, that not everyone can produce verbal self-justifications on demand.) Whether you believe them or not, make sure you are aware of at least one circumstance-based explanation for what they did.

Notice which situations elicit more of the disliked behaviors than others. Everybody has situations that bring out the worst in them, and when the worst is already getting on your nerves, you should avoid as much as possible letting any extra bubble to the surface. If you have influence of any kind over which roles this person plays in your life (or in general), confine them to those in which their worst habits are irrelevant, mitigated, or local advantages of some kind. Do not ask for a ride to the airport from someone who terrifies you with their speeding; don't propose splitting dessert with someone whose selfishness drives you up the wall; don't assign the procrastinator an urgent task. Do ask the speeder to make a quick run to the bank before it closes while you're (ever so inconveniently) stuck at home; do give the selfish person tasks where they work on commission; do give the procrastinator things to do that they'll interpret as ways to put off their other work.

**2. Increase salience of positive traits.**

Don't look at me like that. *There is something*. It's okay to grasp at straws a little to start. You do not have to wait to like someone until you discover the millions of dollars they donate to mitigating existential risk or learn that their pseudonym is the name of your favorite musician. You can like their cool haircut, or the way they phrased that one sentence the other week, or even their shoes. You can appreciate that they've undergone more hardship than you (if they have, but be generous in interpreting "more" when comparing incommensurate difficulties) - even if you don't think they've handled it that well, well, it was hard. You can acknowledge that they are better than you, or than baseline, or than any one person who you already like, at some skill or in some sphere of achievement. You can think they did a good job of picking out their furniture, or loan them halo effect from a relative or friend of theirs who you think is okay. There is *something.*

Learn more about the likable things you have discovered. "Catch them in the act" of showing off one of these fine qualities. As a corollary to the bit above about not putting them in roles that bring out their worst, try to put them in situations where they're at their best. Set them up to succeed, both absolutely and in your eyes. Speak to any available mutual friends about what more there is to like - learn how the person makes friends, what attracts people to them, what people get out of associating with them. Solicit stories about the excellent deeds of the target person. Collect material like you're a biographer terrified of being sued for libel and dreading coming in under page count: you need to know all the nice things there are to know.

It is absolutely essential throughout this process to cultivate *admiration, not jealousy*. Jealousy and resentment are absolutely counterproductive, while admiration and respect - however grudging - are steps in the right direction. Additionally, you are trying to *use* these features of the person. It will not further your goals if you discount their importance in the grand scheme of things. Do not think, "She has such pretty hair, why does she get such pretty hair when she doesn't deserve it since she's such an awful person? Grrr!" Instead, "She has such pretty hair. It's gorgeous to look at and that makes her nice to have around. I wonder if she has time to teach me how to do my hair like that." Or instead of: "Sure, he can speak Latin, but what the hell use is Latin? Does he think we're going to be invaded by legionaries and need him to be a diplomat?" it would be more useful towards the project of liking to think, "Most

people don't have the patience and dedication to learn any second language, and it only makes it harder to pick one where there aren't native speakers available to help teach the finer points.  I bet a lot of effort went into this."

**3. Reap consistency effects.**

Take care to be kind and considerate to the person.  The odds are pretty good that there is something they don't like about *you* (rubbing someone the wrong way is more often bidirectional than not).  If you can figure out what it is, and do less of it - at least around them - you will collect cognitive dissonance that you can use to nudge yourself to like the person.  I mean, otherwise, why would you go to the trouble of not tapping your fingers around them, or making sure to pronounce their complicated name correctly, or remembering what they're allergic to so you can avoid bringing in food suitable for everyone but them?  That's the sort of thing you do when you care how they feel, and if you care how they feel, you must like them at least a little.  (Note failure mode: if you discover that something you do annoys them, and you respond with resentment that they have such an unreasonable preference about such a deeply held part of your identity and how dare they!, you're doing it wrong.  The point isn't to completely make yourself over to be their ideal friend.  You don't have to do everything.  But do *something*.)

Seek to spend time around the person.  This should drop pretty naturally out of the above steps: you need to acquire all this information from *somewhere,* after all.  But seek their opinions on things, especially their areas of expertise and favorite topics; make small talk; ask after their projects, their interests, their loved ones; choose to hang out in rooms they occupy even if you never interact.  (Note failure mode: Don't do this if you can feel yourself hating them more every minute you spend together or if you find it stressful enough to inhibit the above mental exercises.  It is better to do more work on liking them from a distance if you are at this stage, then later move on to seeking to spend time with them.  Also, if you annoy them, don't do anything that could be characterized as pestering them or following them around.)

Try to learn something from the person - by example, if they aren't interested in teaching you, or directly, if they are.  It is possible to learn even from people who don't have significantly better skills than you.  If they tell stories about things they've done, you can learn from their mistakes; if they are worse than you at a skill but use an approach to it that you haven't tried, you can learn how to use it; if nothing else, they know things about *themselves*, and that information is highly useful for the project of liking them, as discussed above.  Put what you know about them into the context of their own perspective.

**Note general failure mode:** It would be fairly easy, using facsimiles of the strategy above, to develop smugness, self-righteousness, arrogance, and other unseemly attitudes.  *Beware* if your inner monologue begins to sound something like "He's gone and broken the sink again, but I'm too good and tolerant to be angry.  It wouldn't do any good to express my displeasure - after all, he can't take criticism, not that I judge him for this, of course.  I'll be sure to put a note on the faucet and call the plumber to cover for his failure to do so, rather than nagging him to do it, as I know he'd fly off the handle if I reminded him - it's just not everyone's gift to accept such things, as it is mine, and as I am doing, right now, with him, by not being upset..."

This monologuer does not *like* the sink-breaker.  This monologuer holds him in contempt, and thinks very highly of herself for keeping this contempt ostensibly private (although it's entirely possible that he can tell anyway).  She tolerates his

company because it would be beneath her not to; she doesn't enjoy having him around because she realizes that he has useful insights on relevant topics or even because he's decorative in some way. *If you don't wind up really, genuinely, sincerely liking the person you set out to like, you are doing it wrong. This is not a credit to your high-mindedness, and thinking it is will not help you win.*

[1] A good time to practice this habit is when in a car. Make up stories about the traffic misbehaviors around you. "The sun is so bright - she may not have seen me." "That car sure looks old! I probably wouldn't handle it even half as well, no wonder it keeps stalling." "He's in a terrible hurry - I wonder if a relative of his is in trouble." "Perhaps she's on her cellphone because she's a doctor, on call - it then would really be more dangerous on net if she *didn't* answer the thing while driving." "He'd pull over if there were any place to do so, but there's no shoulder." Of course any given one of these is probably not true. But they *make sense*, and they are not about how everybody on the road is a maniac! I stress that you are not to *believe* these stories. You are merely to *acknowledge that they are possibilities*, to compensate for the deemphasis of hypotheses like this that the fundamental attribution error will prompt.

# Seven Shiny Stories

It has come to my attention that the contents of the luminosity sequence were too abstract, to the point where explicitly fictional stories illustrating the use of the concepts would be helpful. Accordingly, there follow some such stories.

**1. Words** (an idea from Let There Be Light, in which I advise harvesting priors about yourself from outside feedback)

Maria likes compliments. She *loves* compliments. And when she doesn't get enough of them to suit her, she starts fishing, asking plaintive questions, making doe eyes to draw them out. It's starting to annoy people. Lately, instead of compliments, she's getting barbs and criticism and snappish remarks. It hurts - and it seems to hurt her more than it hurts others when they hear similar things. Maria wants to know what it is about her that would explain all of this. So she starts taking personality tests and looking for different styles of maintaining and thinking about relationships, looking for something that describes her. Eventually, she runs into a concept called "love languages" and realizes at once that she's a "words" person. Her friends aren't trying to hurt her - they don't realize how much she thrives on compliments, or how deeply insults can cut when they're dealing with someone who transmits affection verbally. Armed with this concept, she has a lens through which to interpret patterns of her own behavior; she also has a way to explain herself to her loved ones and get the wordy boosts she needs.

**2. Widgets** (an idea from The ABC's of Luminosity, in which I explain the value of correlating affect, behavior, and circumstance)

Tony's performance at work is suffering. Not every day, but most days, he's too drained and distracted to perform the tasks that go into making widgets. He's in serious danger of falling behind his widget quota and needs to figure out why. Having just read a fascinating and brilliantly written post on Less Wrong about luminosity, he decides to keep track of where he is and what he's doing when he does and doesn't feel the drainedness. After a week, he's got a fairly robust correlation: he feels worst on days when he doesn't eat breakfast, which reliably occurs when he's stayed up too late, hit the snooze button four times, and had to dash out the door. Awkwardly enough, having been distracted all day tends to make him work more slowly at making widgets, which makes him less physically exhausted by the time he gets home and enables him to stay up later. To deal with that, he starts going for long runs on days when his work hasn't been very tiring, and pops melatonin; he easily drops off to sleep when his head hits the pillow at a reasonable hour, gets sounder sleep, scarfs down a bowl of Cheerios, and arrives at the widget factory energized and focused.

**3. Text** (an idea from Lights, Camera, Action!, in which I advocate aggressive and frequent introspection to collect as much data as possible)

Dot reads about an experiment in which the subjects receive phone calls at random times and must tell researchers how happy they feel. Apparently the experiment turned up some really suboptimal patterns of behavior, and Dot's curious about what she'd learn that she could use to improve her life. She gets a friend to arrange delayed text messages to be sent to her phone at intervals supplied by a random number generator, and promises herself that she'll note what she's doing, thinking, and feeling at the moment she receives the text. She soon finds that she doesn't

enjoy watching TV as much as she thinks she does; that it's probably worth the time to cook dinner rather than heating up something in the microwave because it's considerably tastier; that she can't really stand her cubicle neighbor; and that she thinks about her ex more than she'd have ever admitted. These thoughts were usually too fleeting to turn into actions; if she tried to remember them hours later, they'd be folded into some large story in which these momentary emotions were secondary. But treating them as notable data points to be taken into account gives them staying power. Dot starts keeping the TV remote under the book she's reading to remind herself what entertainment is more fulfilling. She buys fewer frozen meals and makes sure she's stocked up on staple ingredients. She agrees to swap cubicles with a co-worker down the hall. There's not all that much she can do about the ex, but at least when her friends ask her if everything's okay between them, she can answer more accurately.

**4. Typing** (an idea from The Spotlight, in which I encourage extracting thoughts into a visible or audible form so as to allow their inspection without introspection)

George is trying to figure out who he is. He's trying really hard. But when he tries to explain his behaviors and thoughts in terms of larger patterns that could answer the question, they inevitably sound suspiciously revisionist and self-serving, like he's conveniently forgetting some parts and artificially inflating others. He thinks he's generous, fun at parties, a great family man, loyal, easygoing. George decides that what he needs to do is catch what he's thinking at the moment he's thinking it, honestly and irrevocably, so he'll have an uncorrupted data set to work with. He fires up a word processor and starts typing, stream of consciousness. For a few paragraphs, it's mostly "here I am, writing what I think" and "this is kind of dumb, I wonder if anything will come of it", but eventually that gets old, and content starts to come out. Soon George has a few minutes of inner monologue written down. He writes the congratulatory things he thinks about himself, but also notes in parentheses the times he's acted contrary to these nice patterns (he took three helpings of cake that one time when there were fewer slices than guests, he spent half of the office celebration on his cellphone instead of participating, he missed his daughter's last birthday, he dropped a friend over a sports rivalry, he blew up when a co-worker reminded him one too many times to finish that spreadsheet). George writes the bad habits and vices he demonstrates, too. Most importantly, he resists the urge to hit backspace, although he freely contradicts himself if there's something he wants to correct. Then he saves the document, squirrels it away in a folder, and waits a week. The following Tuesday, he goes over it like a stranger had written it and notes what he'd think of this stranger, and what he'd advise him to do.

**5. Contradiction** (an idea from Highlights and Shadows, in which I explain endorsement and repudiation of one's thoughts and dispositions)

Penny knows she's not perfect. In fact, some of her traits and projects seem to outright contradict one another, so she really *knows* it. She wants to eat better, but she just *loves* pizza; she's trying to learn anger management, but sometimes people do things that really *are* wrong and it seems only suitable that she be upset with them; she's working on her tendency to nag her boyfriend because she knows it annoys him, but if he can't learn to put the toilet seat down, maybe he deserves to be annoyed. Penny decides to take a serious look at the contradictions and make decisions about which "side" she's on. Eventually, she concludes that if she's honest with herself, a life without pizza seems bleak and unrewarding; she'll make that her official exception to the rule, and work harder to eat better in every other way without the drag on motivation caused by withholding her one favorite food. On reflection,

being angry - even at people who really do wrong things - isn't helping her or them, and so she throws herself into anger management classes with renewed vigor, looking for other, more productive channels to turn her moral evaluation towards. And - clearly - the nagging isn't helping its ostensible cause either. She doesn't endorse that, but she's not going to let her boyfriend's uncivilized behavior slide either. She'll agree to stop nagging when he slips up and hope this inspires him to remember more often.

**6. Community** (an idea from City of Lights, in which I propose dividing yourself into subagents to tackle complex situations)

Billy has the chance to study abroad in Australia for a year, and he's so mixed up about it, he can barely think straight. He can't decide if he wants to go, or why, or how he feels about the idea of missing it. Eventually, he decides this would be far easier if all the different nagging voices and clusters of desire were given names and allowed to talk to each other. He identifies the major relevant sub-agents as "Clingyness", which wants to stay in known surroundings; "Adventurer", which wants to seek new experiences and learn about the world; "Obedience to Advisor", which wants to do what Prof. So-and-So recommends; "Academic", who wants to do whatever will make Billy's résumé more impressive to future readers; and "Fear of Spiders", which would happily go nearly anywhere but the home of the Sydney funnelweb and is probably responsible for Billy's spooky dreams. When these voices have a chance to compete with each other, they expose questionable motivations: for instance, Academic determines that Prof. So-and-So only recommends staying at Billy's home institution because Billy is her research assistant, not because it would further Billy's intellectual growth, which reduces the comparative power of Obedience to Advisor. Adventurer renders Fear of Spiders irrelevant by pointing out that the black widow is native to the United States. Eventually, Academic and Adventurer, in coalition, beat out Clingyness (whom Billy is not strongly inclined to identify with), and Billy buys the ticket to Down Under.

**7. Experiment** (an idea from Lampshading, where I describe how to make changes in oneself by setting oneself up to succeed at operating in accordance with the change, and determining what underlies the disliked behavior)

Eva bursts into tears whenever she has a hard problem to deal with, like a stressful project at work or above-average levels of social drama amongst her friends. This is, of course, completely unproductive - in fact, in the case of drama, it worsens things - and Eva wants to stop it. First, she has to figure out why it happens. Are the tears caused by *sadness?* It turns out not - she can be brought to tears even by things that don't make her sad. The latest project from work was exciting and a great opportunity and it still made her cry. After a little work sorting through lists of things that make her cry, Eva concludes that it's linked to how much pressure she feels to solve the problem: for instance, if she's part of a team that's assigned a project, she's less likely to react this way than if she's operating solo, and if her friends embroiled in drama turn to her for help, she'll wind up tearful more often than if she's just a spectator with no special responsibility. Now she needs to set herself up not to cry. She decides to do this by making sure she has social support in her endeavors: if the boss gives her an assignment, she says to the next employee over, "I should be able to handle this, but if I need help, can I count on you?" That way, she can think of the task as something that isn't entirely on her. When next social drama rears its head, Eva reconceptualizes her part in the solution as finding and voicing the group's existing consensus, rather than personally creating a novel way to make everything better. While this new approach reduces the incidence of stress tears, it doesn't disassemble

the underlying architecture that causes the tendency in the first place.  That's more complicated to address: Eva spends some time thinking about why responsibility is such an emotional thing for her, and looks for ways to duplicate the sense of support she feels when she has help in situations where she doesn't.  Eventually, it is not much of a risk that Eva will cry if presented with a problem to solve.