

# Spiritual Rationality



*groan...*

This book is written in attempt to lay out what I see as systematic mistakes being made at the intersection of "science" and "spirituality," and to describe what landmarks I can see in charting a path forward.

Move on to the [Introduction](#) to read more.

*This book is a work in progress! Many pages are incomplete, only contain notes, or are stubs.*

© 2024-2025 Dhabī Ibn Mūsā.

# Introduction

---

Mystical and contemplative spirituality has in the West come in, out, in again, out once more, and (no really) back into fashion again. This time around it's with a veneer of scientism, with its concomitant confusions. Even the most "cutting edge" practice and theology among Western Buddhists is sloppy on the metaphysics, usually ineffective, and sometimes harmful.

This critique doesn't just apply to Buddhists of course, but I'm poorly equipped to talk about Sufis or Hasidim or the New Age, though I think my arguments do apply in general.

From the other side of the aisle, so to speak, I see those allied explicitly or just ~ *energetically* ~ with "Rationalism" as making mostly the obverse mistakes. This crowd generally relies on a narrow and overly reified metaphysics, and often have a shallow capacity for introspection.

The unholy love-child of these two ideologies is where I find myself, buffeted by, to me, what looks like insanity on both sides. I'd like to build up a framework to bridge the various discourses I've got some toes in, and which can hopefully contribute to meaningful cultural progress, such as that's possible at all.

I'll try to lay out what I think "right answers" can and can't mean in this domain, and I have some ideas about the direction in which we'll find right answers, or come to stop looking for answers to wrong questions. I don't believe I've answered almost any of my own questions in particular, but nonetheless I believe it can be highly generative to say "not this, not that, none of these."

## Claims

I'll make some more concrete claims at the outset:

I think that almost everyone making strong claims (positive or negative) about spirituality is overconfident and usually making metaphysical errors. Often people willfully come to false beliefs about themselves and the world, *or* dismiss clear evidence about the breadth and depth of experience.

Despite what is said in either direction, I think it remains that we don't have dispositive evidence about a large swath of claims about enlightenment, consciousness, or even rebirth—though I'm still quite skeptical of most of them.

I am not aware of frameworks or theories that are adequate for bridging the gap between spiritual and physical phenomena in a way that does justice to either, nor for distinguishing "correct" or "better" lineages or claims. My *suspicion* is that this will boil down to some fundamental metaphysical-epistemological problems, for which there may in principle be no singular, definitive answers. There will however be provisional, humane, pragmatic answers.

Both folk and academic apprehensions of goodness, beauty, purpose, human meaning, etc., seem usually introspectively shallow and confused, as well as *wrong* and harmful. Traditional spiritual systems seem both very much onto something in this respect, and also insane and stuck in ruts in quite varied ways. All of these seem to be contorted variously by cultural trauma and confabulation. The best depictions are indeed often poetic and rousing rather than concrete, but there's so many of these and they're so conflicting that they can't make any concrete prescriptions. My suspicion (and prejudice) is that the "correct" synthesis is going to be quite spacious and allow for a great deal of particularity and idiosyncrasy, but that it will distinctly *exclude* most of the answers floating around these days.

Lastly, I think most spiritual traditions, and unfortunately most cultures in general, propagate memetic, epistemic, and metaphysical harms. Sometimes these are good choices under harsh tradeoffs, but I believe that we've largely obscured what "good" can or ought to mean, and we need cultures of spiritual practice that are able to bear or transform the costs of shedding preference falsification, dissociation, and submission as fundamental cultural and spiritual techniques.

## Influences and inspiration

This book is an attempt at a synthesis of the perspectives of four authors, in a way which is absent in the discourse. These are: [Eliezer Yudkowsky](#), [David Chapman](#), [Rob Burbea](#), and [Mark Lippmann](#). I also owe a debt of some kind to my time with Sanghananda<sup>1</sup>.

Most of all, I have huge appreciation for Mark. None of this would be possible without his work, and in some sense I see my book as an introduction, or a "popularization" of his work.

## Footnotes

1. IYKYK ↩

# How to read this book / What's up with this website?

---

While this site presents the pages in a linear order, it wasn't really possible to structure the material fully linearly, so some pages link forward or backward. Still, it's meant to be read in the order presented, as I make distinctions that build on one another.

## **This project is a work in progress!**

Most pages are still incomplete and will show a callout like the one below, expressing various states of completeness. Those without callouts are no longer being drafted, but may still change in the future.

*On the pdf, these are not shown.*

*This page is incomplete or in a draft state.*

## **Jargon / technical sections**

Some explanations begged to be explained with reference to a specific technical model, or at least riffing off of one, but I also wanted this material to remain mostly accessible. In addition, unfortunately, "technical" is often with reference to some eclectic mix of disciplines and traditions, so I expect most people to find at least some of these unhelpful. Therefore, these sections are presented with a choice of a more accessible explanation, and separately a more technical or just abstruse explanation.

*On the pdf, these display one after another, and are not interactive.*

**Vernacular**

You can click on the other tab to view a more in-depth (but jargon-laden) explanation. Your choice will be saved for sections like this within this site.

### **Technical**

Choices are persisted in localStorage; in addition they are persisted in the URL's query string, so that you can link to a page with a specific jargon level.

### **Expandable sections**

There are also expandable sections like this one, for digressions or elaborations that aren't vital to the main text.

*On the pdf, these display expanded, and are not interactive.*

**Click here for some tasty details...**

(Or not so much, in this case.)

# Metaphysics

---

*In this section:*

*Metaphysics* is these days an unfortunate word, but in some sense it remains the most fundamental subject to all inquiry:

- [What, and Why, Metaphysics?](#)

Metaphysics is of particular interest in trying to understand phenomenology and spiritual experiences. I discuss some aspects of phenomenological metaphysics:

- [Ways of Seeing](#)
- [Reification](#)

I then discuss properties of many metaphysical systems, and certain errors in those views:

- [Eternalism, Nihilism, and Nebulosity](#)
- [Good, Right, & Wrong](#)
- [\(Mis\)use of Language and Rhetoric, re Rightness](#)
- ["Real"](#)

Lastly, people have many different beliefs about metaphysics, and it's quite unclear how to determine what's right, or to come to consensus:

- [Metaphysical Disagreement and Determination](#)

# What, and Why, Metaphysics?

---

This isn't abstract, because this affects the whole texture of your sense of existence. The whole sense of meaningfulness in your life, the whole vision and feel of what your life, existence, and practice is depends on what you believe is real and or [sic] not, and what's not real. How are we going to know? How are we going to decide?<sup>1</sup>

There's some trouble here because the word 'metaphysics' is too often these days basically used by cranks, and in any case there isn't a standard definition, but the definition I'll use I think covers most of the modern usages pretty well. Anyway, it's not definitive and I'm holding it lightly.

I'm going to define metaphysics as the combination of ontology, cosmology, axiology, and epistemology.

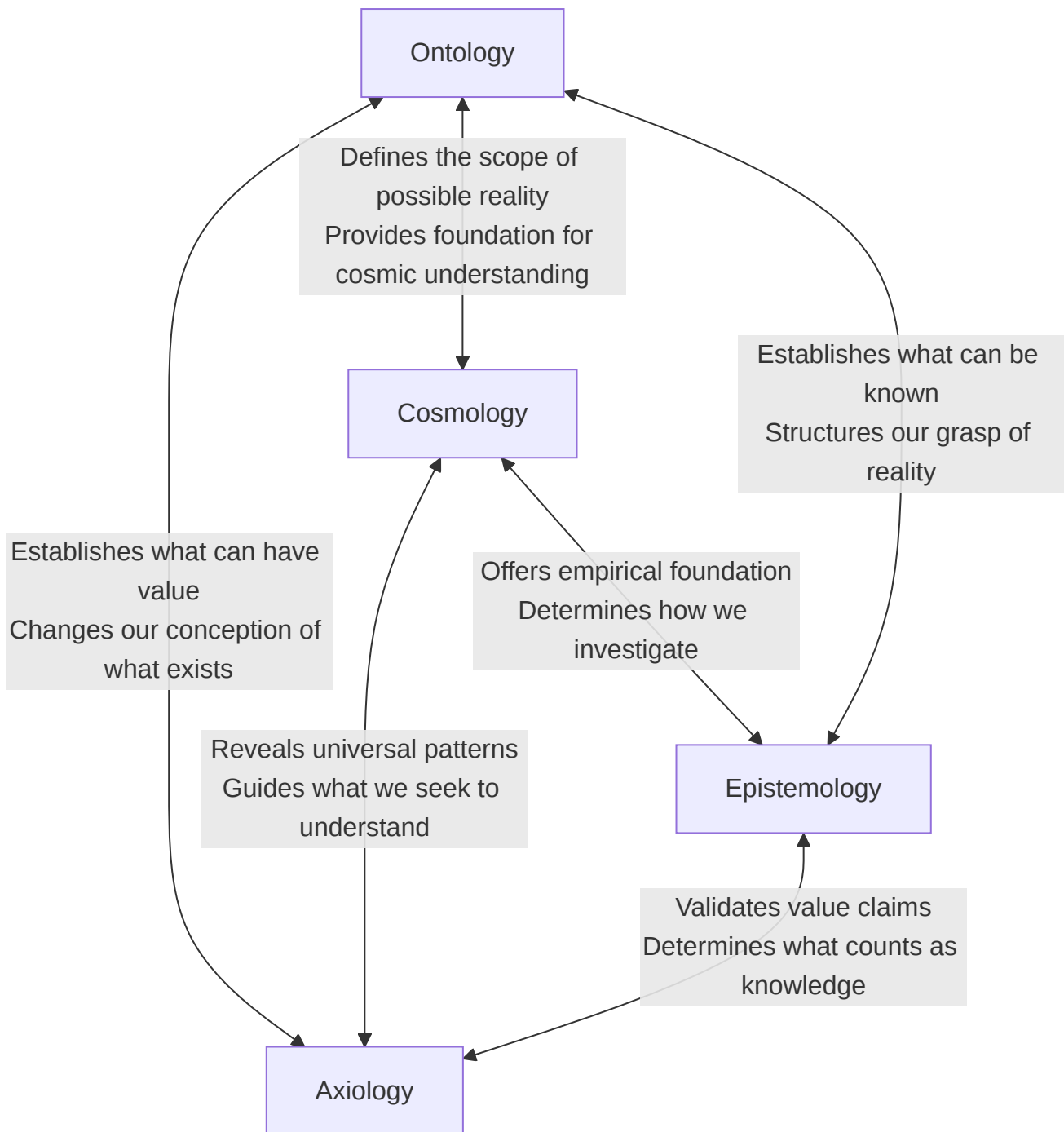
**Ontology:** what exists, and what it means for something to exist.

**Cosmology:** what is out there, including in the far reaches of time and space.

**Axiology:** what is good and right, what matters, and how important this or that is.

**Epistemology:** how we can know.

Clearly, these concepts are closely related:



I asked Claude to write a diagram of the relationships between each and it did a great job, actually.

**One way of conceiving of these is that they are the fundamental human hyperpriors.**

The point here being, we have priors over what kinds of evidence we expect to observe, top to bottom, and how we plan to (or, implicitly, naively, will) update, with respect to our world-models and our goals and values.

I'll give some illustrative examples to try to make this concept feel more tangible.

In various systems of Christian metaphysics, we get something like this:

- Ontology: souls, God, magic (or at least, whatever is responsible for miracles)
- Cosmology: Heaven and Hell, sometimes also Purgatory; time beginning from Genesis



- Axiology: obeying God's will, salvation as the highest good, virtues like faith and charity; Sometimes also God's will as *defining* the Good
- Epistemology: scriptural and church authority, divine revelation through scripture and prayer

In some sense this is a bit of a caricature, or at least an exaggerated description. In practice I don't think almost anyone actually uses solely this metaphysics. I expect that, even in the traditions that teach it, eternal damnation is not that salient on most days of people's lives, for all but the most devout. Similarly, the demands of family and community, let alone ordinary personal satisfaction, are probably not constantly being practiced as devotion to God, etc. etc.

To be clear, in general, one's explicitly endorsed metaphysics isn't necessarily all that's going on, and in practice I think *everyone* is using a messy collection of competing perspectives.

(In particular also, of course, "Christianity" isn't one thing, etc.)

Let's take rationalism. Again not to caricature, as rats are a diverse bunch, but these are pretty common features:

- Ontology: physicalism, computation as the basis of consciousness, sometimes a kind of mathematical platonism by way of Tegmark
- Cosmology: many-worlds interpretation of quantum physics, deep time both forward and backward, potential simulation hierarchies
- Axiology: consequentialism, usually utilitarianism, often longtermism and certain kinds of techno-optimism, sometimes priority on reducing suffering
- Epistemology: Bayesianism, reductionism

Interestingly I think rationalists tend to be quite a bit more "devout" than probably 98th percentile Christians in the US, and I know many people who are actually just whole-hog on basically the picture I drew above.

This one is faintly less fair, but here's an attempt to describe the standard coastal liberal worldview, perhaps as of 15 years ago:

- Ontology: Material reality is primary, consciousness emerges from brains, social constructs are "real but not really real," abstract concepts like rights and democracy have a kind of quasi-reality
- Cosmology: Big Bang, evolution, progress as a general historical pattern. Often has a kind of vague, lossy scientific picture
- Axiology: Individual autonomy, harm reduction, fairness/equality, tolerance of difference (except intolerance), personal growth/development
- Epistemology: again "science," but more in a social/institutional sense, trusting experts and consensus rather than direct engagement with method. Strong separation between "subjective" domains where personal experience is authoritative and "objective" ones where it isn't.

Ordinarily we might call all of these "worldviews," which is fair enough, but a large part of the point of calling them each a metaphysics is to note a. how they're all fairly tightly interwoven, each aspect buffeting what kinds of inferences can be made with respect to the others, and b. that they're often *totalizing* in a sense. Even from the contemporary liberal view, they can regard contrary perspectives from afar (especially those from a [fargroup](#)), but it's really quite difficult to deeply engage with ideas that majorly challenges these principles.

Yes yes, before you call me snooty or a hypocrite, I have a metaphysics too, and I'm absolutely vulnerable to this totalizing quality. That's really the point here, and that's why it's so important. It's somewhere between exceedingly difficult and impossible to escape to clutches of metaphysics in general, and I need to be able to talk about what properties a metaphysics has, and what consequences that has for inferences made under it.

The point of discussing metaphysics is not to sneer at people for believing things, at the very least not *for believing things whatsoever*. Mostly, I mean to say: these systems of beliefs exist, indeed they have such and such properties, and they affect perception and cognition.

Let me also say definitively: the details of one's metaphysics **matter**. *Surely*, it's of great consequence what is valuable, *surely*, it's of great consequence what's real, *surely*, etc. etc., complete the litany however you like: physical death, heat death, childrearing, purpose, rebirth, free will, love, quantum many-worlds, simulation hypothesis, moral truth, other minds, time's arrow, on and on.

To some extent this is foreign from the modern secular malaise of hardly-even-believing-anything, but even from such a perspective it should be clear that *if you actually believed* in some distinct metaphysics, then yes! that would have really substantial consequences on how you lived your life, how you related to others, and what would appear important and real and worth investing in. I should also say, the modern secular view nonetheless involves a metaphysics, even if it's mostly invisible to us because it's assumed so broadly everywhere now.

Perhaps, and this question will weave in and out of the whole book, for some of these questions there doesn't exist, and in principle *can't* exist a "right" answer. Often these questions are deeply personal and idiosyncratic, and in principle a person, or a mind in general, could choose one thing or another without contradiction. See [Good, Right, & Wrong](#) for more on this question.

### **Secular western metaphysics is taken for granted, but is very modern, and often cartoonish**

There are two points here:

1. Many of the details of the default secular worldview are extremely modern. Included in these would be things like the social institution of science, specific models like evolutionary biology, the Big Bang, and current estimates for the age of the universe ( $\sim 1.4 \cdot 10^{10}$  years), and human rights, which all came to consensus quite recently, some of these less than a century ago. To be clear, *modern* doesn't unreliable or faddish, but simply recent in history. These ideas which we take for granted *were simply unknown* to past generations in Western culture, let alone in world culture cross-anthropologically.
  - I think there's something pernicious with all of these about consensus shorn of historical context.
  - Perhaps, we could say, "this perspective is one that a human mind can experience (as indeed I am experiencing it) and the view has consequences in terms of the vague background intuitions I have, what seems salient to me, and what kinds of inferences I'm likely to make."
  - Some very important questions are still uncertain even in cosmology, eg. heat death.
2. The Big Bang is probably a cartoon story in your mind. If you're not at least a highly competent amateur physicist, I'm going to bet that you understand very little about modern scientific cosmology. Not only do you likely not understand the model, you also probably have no idea how the consensus came about, what contrary theories or data exist, etc. etc.

- To be clear, this *definitely* includes me.
- Same goes for evolution, which is indeed mostly not understood even by the "general educated public."
- This is also ignoring that humans mostly can't perceive magnitudes like  $10^{10}$ .
- Included in this would even just be "science," which I think in the popular imagination and discourse amounts to not much more than a priestly class who perform magical rituals which produce knowledge by basically inscrutable means.

## The thicket of views

We can analyze metaphysics from a propositional, epistemic, or phenomenological angle. I think the phenomenological level ends up being the most revealing, and I think the propositional is where we can easily get caught and make no progress at all.

"Vaccha, the position that 'the cosmos is eternal' is a thicket of views, a wilderness of views, a contortion of views, a writhing of views, a fetter of views. ..." <sup>2</sup>

"Whereas some recluses and brahmins, while living on the food offered by the faithful, engage in wrangling argumentation, (saying to one another): 'You don't understand this doctrine and discipline. I am the one who understands this doctrine and discipline.' ..." <sup>3</sup>

While I hardly take an orthodox Buddhist angle here, this is clearly an ancient problem. I'm also not clear that I have a solution to this problem, but I'm glad to say that strictly staying at the propositional level is seldom generative. That's part of the point of this whole book: it's not clear in principle how to [make updates](#) or come to consensus about metaphysics! (And, in practice, this kind of discourse usually gets nowhere.) Indeed, this is a very familiar kind of doctrinal bickering between schools.

## No universally compelling arguments

Following Eliezer, I agree that there are no universally compelling arguments <sup>4</sup>. This is to say, in principle if one mind thinks or perceives one thing in a specific context, another could simply "choose" something else, or the opposite. This is not to say that therefore nothing is real or true; we humans will continue to ~all agree that 2 is bigger than 1, and that on earth if you drop a marble from a building it will fall, and these models and predictions seem to continue to be consistent and predictive.

However, among humans there is a riot of differing metaphysical perspectives, and it's not clear that consensus is possible:

### Vernacular

There are so many different contradictory perspectives about what's true and real, that I suspect it might be *in principle* impossible to reach consensus. I suspect that human minds are flexible enough, and that there won't be a stable basis on which we can construct a "correct" answer for most of the content of metaphysics in general.

(Cosmology is mostly only an exception here if you accept a ~"scientific" epistemology, which is also not guaranteed.)

There's so much diversity in what is valorized that I don't think that there's going to be an inescapably "correct" answer, in the way that we find for mathematical or empirical questions, again my suspicion is that human minds are simply too flexible here. Largely the same goes for ontology as well.

Regarding the flexibility of value, I'll discuss all of this in more detail in the section on [axiology](#).

### **Jargon**

My *suspicion* is that metaphysics is going to be underdetermined. In some ways this is a challenge for my whole project, though mostly what I'll argue for, in the rest of the section on metaphysics, is that we don't have a "correct" metaphysics, and that likely there in principle cannot be one.

Some parts of the metaphysics will ground out to ~empirical fact, eg. most of the details of the cosmology, but it's not obvious to me that the occamian prior is "inescapable" for a human mind.

More relevantly, there's so much diversity in what is valorized that I don't think that there's going to be an inescapably "correct" answer, in the way that we find for mathematical or empirical questions, again my suspicion is that human minds are simply too flexible here.

I feel pretty confident that the basic biologically determined ontological priors are adapted to get human-sized objects mostly right without too much superstition, but that we don't come baked in with stable intuitions about whether a machine can be a person, or what it means for something to be "dirty," or even, classically, whether an object always remains the same one over time. The following few sections will discuss this problem, among others. See especially "[Real](#)".

I'll discuss questions of value and culture in more detail in the section on [axiology](#).

A lot of public discourse actually happens between people who have incompatible systems of metaphysics. Seen this way, it's *obvious* that they can't make any progress: perhaps, because they don't have means to make changes to their metaphysics, but at least because they're trying to argue over claims that are downstream of their metaphysics, without surfacing their underlying metaphysical differences.

## **Epistemic consequences of metaphysics**

More productively, we can notice that a person's metaphysics will affect what inferences they make. This is both subtle and overt. Of course, rationalists might have explicit models about evolutionary psychology, but there's often also a tendency to see all interactions in terms of some imagined notion of the incentives of on some prehistoric evolutionary african savannah. We might imagine here also beliefs about literal miracles, versus faint inklings of providence and the hand of God.

Metaphysical beliefs are the greatest of the geological forces shaping the so-called "salience landscape." I would say here, we can never do a totally neutral search over the space of hypotheses, we *start* searching for an explanation with some opaque, implicit assumptions about what kinds of explanations might exist, and where to look for them.

## **Phenomenological metaphysics**

Firstly, just to clarify: phenomenological metaphysics could probably mean either "metaphysics *with respect to phenomenology*" (as in, what is consciousness, qualia-based theories of value, etc.), or "the

phenomenology of metaphysics," and I'll be talking about the latter here, though the former will come up again in the section on spiritual practice.

Perhaps more than on one's epistemics, one's metaphysics has a huge part in shaping the contents of one's phenomenology.

Ordinarily, one doesn't see perceptions, they rather constitute what we experience as real. As we experience them, they are usually *out there*, rather than *in here*. I could say that we see *through* perception, or at least *with* perception. It's not even that, propositionally, we say "ah well that object which I hypothesize to exist in some 'reality', based on its appearance to me, I would categorize it as x"—a sentence no one has ever said in earnest—rather indeed we *experience it* as a member of that category. That simply *is* a dog, or worse, that sex act simply *is* disgusting, etc. Sometimes a perception is uncertain, but the *quality of being uncertain* is usually clear as day.

My impression is that children basically know this, at least about other people. Alice sees x this way, Bob sees it that, they both experience their perception as utterly real, and I see it yet differently. "Obviously, their perceptions are just perceptions" (and then perhaps) "whereas mine are *actually* real."

Borrowing from the language of Internal Family Systems, I'll say that perceptions are more "blended" the less easily we can take them as object or try on another perspective. Usually, our basic metaphysical assumptions are so blended that they're simply stuck to our faces, to mix metaphors. We often usually *can't even see them as assumptions*, even if we have some flexibility with perceptions of eg. beauty or disgust. "I have a body", "I am looking out of my eyes", "the ceiling is up", or even "this person is separate from me," these are so real and so ever-present that they're somehow both invisible and utterly stable and solid.

One's metaphysics shapes "the very being and seeming of the world." Evocatively, we can compare here the standard image someone in the throes of mania, seeing CIA agents or angels around every corner. Even in ordinary, non-pathological cases this can nonetheless be quite stark. An ok example, though this also isn't *quite* metaphysical, would be reports of travelling from the dead of winter to a tropical country, that life and experience seem to open up entirely, that the possibilities for connection and flourishing become obvious where before they seemed only imaginary.

It's been difficult to situate rich humane meaning in modern scientific cosmologies, and indeed we can say that there's some attraction to placing humanity, which most people regard as being the locus of meaning and goodness in the universe, as the *cosmological* center, in countless creation myths through Galileo, and still to this day in many fundamentalist religions.

This kind of "existential seeming" is where spirituality becomes relevant. This is also where a lot of "woo" comes in, though unfortunately often through "hyperstitional" views and practices, which I think are mostly harmful.

## Footnotes

1. Rob Burbea, 2014. [Questioning Awakening](#) ↩
2. [MN 72](#), trans. Thānissaro 1997 ↩
3. [DN 2](#) trans. Bodhi 2010 ↩
4. Eliezer Yudkowsky, 2008. [No Universally Compelling Arguments](#) ↩

# Ways of Seeing

---

A *way of seeing* is an orientation or a pattern of relating (whether to a physical object, a person, a symbol, etc.), which gives rise to some cluster of emotions and perceptions. This is one of the basic concepts in how Rob Burbea taught and explained meditation, which is where I get it from.

That definition is quite abstract, so I'll try to make it more concrete.

Reading a math textbook, one might notice any of these, or any of these might be salient at one time:

- Mathematical arguments
- Relationship to other mathematical theories
- Typography and design
- Physical presence as an object
- Context within the author's life

These are all pretty mundane, but any of them define a pattern of salience and attention, as well as emotional and subtler perceptual qualities.

Religious and sacred objects, for those steeped in a given tradition, will arouse various emotional, perceptual, and behavioral qualities. These might be static or dynamic: here I'm thinking of the way the Torah scroll is taken out of the ark and held in the air, and presented to the congregation who sing for it. Other traditions have many different patterns of embodied, dynamic relations of sacredness with particular objects, or with the architecture of a temple, etc., which again embody and arouse a way of seeing which is practiced with respect to that object.

Sometimes ways of seeing can be aroused with respect to inert symbols:

- The swastika
- Pop culture icons
- Sexual/taboo symbols, eg. the Ring of O

Ways of seeing might be embodied in personal relationships:

- How a parent sees their child
- How someone views a loved one's grave
- Relationship with an enemy or abuser

Often ways of seeing reflect an experience or perception about one's place in the universe, or one's felt sense of status and power. The phrases below each express a relational quality between self and world, particularly the social world:

- "I'm on top of the world"
- "I'm at rock bottom"
- "The world is against me"

- "I'm on another level"

Even plainness or emotional neutrality is a way of seeing, at least of a sort.

---

Note, though we might describe a way of seeing by a phrase in English, ways of seeing are generally non-propositional, and they're not necessarily about specific thoughts one might have. In fact, a way of seeing often has nothing to do with words at all.

A way of seeing is a particular perspective, stance, or position in relationship which brings out or arouses some shape of meaning with respect to an object, situation, person, etc. Ways of seeing often work by amplifying the salience or emotional resonance of certain parts of experience over others. In any of the above examples, there's a detail and richness of meaning which is more or less the *manifestation* of the way of seeing at play. Phenomenologically, this might include feelings in the body, images, perception of situational awareness, or felt qualities of affordance for movement or body positioning, to name a few qualities.

**I claim that perception and "belief" or "ideation" are much more contiguous than people usually imagine**

This is hard to argue for definitively, though I expect my meditator friends are mostly nodding their heads at this section. You can believe me or not, but here's at least some gesture at what I mean by this:

Taking visual perception in particular, I think there's something much more subtle going on than we usually think. I claim that "what something looks like" is often ~"mostly" determined by what we believe about it, than its actual physical structure. We have of course the adage, "beauty is in the eye of the beholder," etc.

My impression from my own meditation practice, and reports from some meditation teachers, is that "mental imagery" is often more subtle than we usually notice, and more like contiguous with ordinary visual perception. There's a spectrum here. Near one end might be purposefully overlaying perceptions onto "external" visual perception, eg. "seeing" how a chair would look in a room that one is decoating, while looking around with the eyes still open. Perhaps in the middle would be something like a rock climber "seeing" routes. On the far end, I've heard claims from meditation teachers to the effect that, with the eyes closed, if one doesn't construe one's body as a body-shape, that body sensations will be experienced by default as occurring on the surface of a sphere. (The interpretation here being that proprioception is more like visual than somatic, despite being so *apparently* somatic.)

Visual perception here is just one example. Most of what I want to say here is that I think in every domain of perception there's a much more dynamic relationship between perspective or stance, and apparent perception, than we usually notice or think possible. I'll claim that there's quite a lot of "give" in perception in general, and that this is exactly where ways of seeing function and are manifest.

There will be more discussion about this in the [section on practice](#), but I'll touch on it briefly here. There is a huge swath of practices which involve giving rise to particular emotions and and perceptions. Among these:

- "Prayer"

- (This evokes some Hollywood image but in my mind is much more general)
- Deity Yoga
  - Briefly, this is a cluster of practices which involve visualizing a deity, sometimes in front of oneself, or sometimes with oneself as the deity
- Many meditation practices
  - Many feature cultivating "equanimity," or sometimes cultivating a perception of un-reality or weakened importance for objects, emotions, etc.
- A large swath of western "therapy" practices
  - This can range from eg. analyzing beliefs and feelings and practicing new ones (a la CBT), to perhaps on the other end, visualizing perfect parents and "taking on" their care as a new emotional basis (IPF)



# Reification

---

The [OED entry](#) for 'reify':

transitive. *To make (something abstract) more concrete or real; to regard or treat (an idea, concept, etc.) as if having material existence.*

Here's about how I use the word:

*To make a perception, feeling, belief, or conception more real, solid, or salient, by any means.*

I would say perception and belief are often downstream of our metaphysics. The definition from the OED describes one common kind of reification, which I'll discuss in more detail [later](#).

I'd also distinguish between *propositional* reification and *phenomenological* reification, which of course are often intertwined.

Propositional reification often takes the form of constructing a model or principle to be taken as the basis for reasoning or acting with respect to some domain. Some examples:

- Social contract theory
- "The scientific method"
- Utilitarianism
- Scriptural inerrancy
- Medical diagnostic criteria
- Design frameworks
- The divine right of kings, or similarly in China, the mandate of heaven

Note, sometimes these can be used manipulatively, but often they're clarifying and generative. A model can be a useful set of assumptions for thinking about a messier domain. Furthermore the concept is even more generic; we might say that physics is "highly reified," (in the sense that it claims that all physical phenomena are constructed out of an exact list of simpler ones) but this is not to say that it's necessarily mistaken, confused, or harmful. On the other hand, reification is more important to talk about when it's being made *inappropriately* as a cognitive or discursive move, and so I'm more likely to talk about cases that are problematic or contentious.

In some sense, any concept at least a little bit reified. Even just to bring a concept to consciousness enough *to consider it at all explicitly* is to reify it, among the kaleidoscope of ways that we could see something. This is great, this is helpful, this is how thought and perception work whatsoever, but also reification is a quality that can lead us to become confused.

Reification is probably necessarily "reductive", that is, it takes something complex and reduces the apparent complexity. If we take *this* theory or that perception as more real than that another, then whatever complexity which is captured by the second is at least less salient, if not eclipsed entirely.

(I unfortunately waffle between this sense of reification as "taking a reductive view" and the one at the top, of "making a perception more apparently real", and so might be a bit unclear at times. I usually want to refer to this whole broad cluster of motions, without needing to be too precise.)

Phenomenological reification overlaps with propositional reification in many of the specific examples. Still, we can usefully highlight the experiential, perceptual, and emotional aspects.



One nice example is the classic old crone/young woman image. Some people will get "stuck" in one perception, but many people can actively switch between them. At one time, generally people can't see *both*, we can only perceive one; while we perceive the old woman, she *becomes real*. Obviously, both perceptions are something like "latent," and neither can be *more real* than the other, but we've *reified* one. (Fwiw, I think this point still applies even if you can see both at the same time.)

This is sort of an easy example, but what's more interesting is all the stuff that's so much stickier. Ordinarily we take a feeling, or a [way of seeing](#) as straightforwardly real, leaving out all the other ways we could see or feel about something.

## (In)appropriate (De)reification

	appropriate	inappropriate
reification		
dereification		

So there's a 2x2 here, and I'll gesture at what's in the different boxes.

### Inappropriate reification

"Inappropriate reification" I get mostly from Mark Lippmann, and is a distinction I haven't seen much elsewhere, though it rhymes with some concepts in Zen. (Yes, sorry, and not even pop Zen either.)

Mark often uses the concept to refer to someone prematurely, or perniciously, taking something as real or stable which "should" have just been provisional. Often a person might notice that x seems to cause y, and become convinced that x always causes y, or that if y is true then x must be, etc. Some concrete examples:

- "If I sit in this posture, then I'll get concentrated in meditation"
- "Since strength training fixed my knee pain, all joint problems must be due to muscle weakness" (I don't even think this is a strawman, I expect that some people believe things like this)
- "I feel ashamed as a person because of what my mother said to me as a child"

Eg. in many examples from the first list of propositional reification, often inappropriate reifications take the form of reifying a particular perspective as absolutely, or *exclusively* real, when in fact it's only a helpful set of assumptions, or it's oversimplified.

Sometimes I also want to say that something is "overreified," meaning something like that it's been taken as "more real than is appropriate" (but not to say that it's not real or important at all).

## Appropriate reification

Like I said, reification is helpful, and even basic human functioning relies on it, I'm thinking here say of object permanence. I would say that phenomenally constructing stable distinctions is probably the basis of perception whatsoever, and so is certainly good, useful, appropriate, etc.

Unfortunately, however, just about everything beyond that is going to be contentious. I'm thinking here of big notions of Justice, Truth, Beauty etc., maybe also one's "Path" or "Destiny." Even things like cleanliness, politeness, and correctness are non-obvious, at least at the margin.

## Dereification, in general

A very common form of dereification is literally telling someone "that's not real". Some dereification occurs ""naturally"", likely everyone is familiar with waking up from a dream and the reality of the dream dissolving as they come to. Often, there's learning and slow decay that happens naturally, as one way of seeing something replaces another.

Dereification is a important aspect of traditional Buddhist models, and maybe marginally in some other traditions. Sometimes, dereification is even *quite reified*, and seen as the goal of the spiritual path! Reports of meditation-induced dereification often include descriptions like "cutting through" or "teasing apart" perception and phenomena in general, either resulting a phenomenon appearing as flimsy or thin, or in its ceasing altogether, up to and including body, mind, and world.

## (In)appropriate dereification?

Obviously, these will also be pretty contentious. Many people (likely most reading this book) think that God is not real, and should not be taken as real, but would be aghast at the idea of regarding the love for a child as "not real." I won't say more about this here, much of what I'll discuss in the sections on [practice](#) and [axiology](#) will be about this question.

# Eternalism, Nihilism, and Nebulosity

---

Eternalism and nihilism are two poles of a tendency in relationship to either domains of meaning or systems of meaning. Eternalism seeks to stabilize and [reify](#) a model or [way of seeing](#), whereas nihilism seeks to destabilize and dereify. I get these concepts from David Chapman; they come originally from at least a specific lineage of Tibetan Buddhism, though I suspect his interpretation is a bit modern and idiosyncratic.

*Eternalist* views see some domain of meaning or reality as stable, fixed, undeniable, or singular. Some examples:

- Justice, as determined by eg. divine command, or "natural law".
- Positivism tries to construct logic as an undeniable and flawless basis for all knowledge
- *Soulmates* as a feature of many folk fantasies or perhaps even "theories" of romantic love
- *Purpose*, as a story for the goal and value of a particular human life, eg. "I was put on this earth for a purpose"
- Souls as the basis and substance of human personality and identity

Unfortunately I sound pretty unsympathetic here, but it's hard to treat these sympathetically because (from my perspective) they're so obviously confused, or just empirically or even logically false.

*Nihilist* views construct a domain of meaning as unreal, insubstantial, or irrelevant:

- Atheism, generically, both denies *cosmologically* the existence of God, but also usually regards divinity, ritual, and sacredness in general, as meaningless or harmful.
- Views such as that one's legacy is meaningless, because in the future one will be dead, so therefore one's legacy is "not real" or "not meaningful" to oneself.
- Many formulations of moral relativism maintain that moral beliefs are of no consequence with respect to one another, and often that morality itself is basically meaningless or inconsequential.
- Similarly, some forms of epistemic relativism effectively deny that "knowledge" can exist whatsoever.
- Views to the effect that human meaning and value are actually meaningless because they are believed to be based in biological processes

While all of these examples are given *propositionally* because this is a written medium, the more interesting aspect is something like *perceptual* or *existential*. These kinds of views, most of them more or less [ways of seeing](#), have intuitive, phenomenological effects, besides whatever arguments or ideologies they cause us to deploy or ally with.

Also, to clarify, 'eternalism' doesn't necessarily refer to stability in time, rather it means something more like "utterly fixed in meaning." Similarly, 'nihilism' doesn't here just refer to the view that *everything* is meaningless and pointless, eg. "existential nihilism," but more to views that specific things are.

Nihilist views say something like, "I shall not care about this," whereas eternalist views say "I shall care unceasingly about this." They both try to admit no wavering in the perception of meaning. Eternalist views

are often about constructing stable, unambiguous *justifications*, whereas nihilist views often say, "there can be constructed no stable justification for this, therefore you get nothing at all."

## Nebulosity

Eternalism and Nihilism, in this model, are ways of creating stability in the face of *nebulosity*. Chapman gives two standard examples to illustrate what nebulosity means:

- "Nebulous," literally "cloud-like": clouds are a real phenomenon, but blend indistinguishably into each other and into the background sky. Up close, or *inside* a cloud, like in an airplane, it's unclear if the cloud even exists, and it has no apparent boundary at all.<sup>1</sup>
- How many pebbles are in a bucket? Well it depends on what counts as a pebble, since there's probably some grit and sand in there as well, and of course grit and pebbles are really the "same" thing, just of different sizes. There isn't quite a "correct answer" except for a specific purpose.<sup>2</sup>

Those are excellent examples but I want to stress that nebulosity is much more general. Often we're talking about ontological (and eventually axiological) nebulosity. For example, even physical quantities which seem quite certain and definite are not necessarily so simple, and sometimes were unknown until recently in human history:

- Mass is generally understood to be definite and clear, up to the accuracy of the measurement. This is correct, but if you put two objects in someone's hand that are similar in mass but different in density, and ask which is heavier, most people will pretty consistently "feel" the denser object to be heavier. If I recall correctly, this is true up to substantial differences in mass, holding the difference in density steady.
  - This is also ignoring that units of "weight" are defined as a force exerted on an object by a particular gravitational field, rather than an intrinsic property (mass). People say, everything is "lighter" in the freefall of an orbiting satellite—wait what? I thought weight was definite? etc.
  - Mass is a fine concept, and often more useful than what we now regard as either derived or compound units like weight and density, but it's not obvious that human-scale perceived weight "needs" to be the same as mass.
- Temperature, up to around the 18th century, was taken to be too confusing and variable to be reduced to a consistent measurable quantity. Perceived temperature is also a function of wind chill, humidity, air pressure, and the person's physiological state, even though we can now measure temperature quite accurately.<sup>3</sup>
  - Indeed, "how hot is it out" has much more to do with perceived temperature than physical temperature.
  - Even things like "boiling point" and "freezing point" are actually fairly nebulous. Water will evaporate *faster* at temperatures well below boiling point at a given pressure, and very clean vessels which provide no nucleation points can avoid approaching a rolling boil well above boiling point. Similarly for the freezing point, highly purified water in a very clean vessel can stay liquid decently far "below freezing."

Concepts like the borders of a country, GDP, unemployment rates, and of course race, are much more nebulous than temperature and mass. This is of course letting alone those like "freedom," "love," or "democracy."

I would claim that nebulosity is much closer to the fact-of-the-matter of human concepts and meaning, namely that they are substantive, meaningful, and worthwhile, while also ambiguous, uncertain, and seldom sharply defined. I would claim that all of the examples above of eternalist and nihilist views are at least partly true or real, or true in some sense, but a. usually not in an exclusive sense, and b. in a sense which blends indistinguishably into other, apparently "contradictory" perspectives.

To elaborate a bit on some of these, at least specifically from my own perspective:

- I believe there are, relatively speaking, justice and injustice, and that they are meaningful and important, but I don't believe that they are "mind-external" in the way that physics is, and I don't believe that they were given or defined by an intelligent creator. I also think justice is substantially culturally bound, even though I still feel strongly about notions of justice which other cultures, present and past, would disagree with.
- I believe there is profound intimacy and "real," sometimes earth-shattering, romantic love, but that it's contingent on circumstances and personal history, and that it's limited in time, at least to natural lifetimes but perhaps much shorter, ie. that perhaps romantic love is both "real," and "not meant to last" past the length of time needed to raise children, or even less.
- I don't believe in a cosmological God, but I believe that divinity and holiness are "real" in a sense such that throwing them out is something like "wrong" or "mistaken," and that it's often "correct" to construe something like God, at least in some culturally specific context.

The point here is robust agnostic of these specific cases. I'm trying to describe structures which might we might want to, or do often, describe as though fixed and definite, and which I think are in fact unstable and nebulous, while still "real" or substantial.

## Nebulosity and comprehensiveness

I want to pick on a specific confusion I see all over the place, in philosophy in general, and especially among rationalists.

Models are *lossy*, or as people often (mostly wrongly) repeat, "wrong." Many models are presented by their authors, explicitly, as comprising some useful heuristics, or some set of generative assumptions; these can however be *mistakenly* taken as comprehensive and authoritative by someone who isn't paying enough attention. (See [inappropriate reification](#).)

That's a boring case but still a common failure mode. Examples that come to mind are eg. psychiatric diagnosis again (where, specifically as described by the DSM, the diagnostic categories are not regarded as necessarily forming natural kinds), perhaps also speciation in both biology and linguistics. Even concepts like covalent vs. ionic bonds are actually continuous, on the same dimension! In these cases the model is presented as lossy but is still predictive in a range of contexts.

More interesting are cases where it's ambiguous or uncertain whether a model is comprehensive. My understanding is that the modern posture taken in physics is something like, "(because of the failure to yet synthesize QM and relativity) indeed, something about these models is wrong and we're pretty confused, nonetheless they're extremely predictive in the domains in which they're regarded as predictive, and so you should absolutely not expect to be able to achieve physical effects contrary to the existing models." (Perhaps with some exotic theorized exceptions like vacuum collapse, etc.)

What's remarkable *metaphysically* and culturally about physics is that it's referred to outside of academia as singularly authoritative and comprehensive, which is not even the posture taken *by actual physicists*. My point here is not to take a jab at physics, or to claim that some metaphysical gymnastics will let you perform miracles—I have no such expectation. My main gripe here will be with respect to theories of value, but I think the point applies for all of metaphysics:

If you have a model or a theory, at what point can you say, "indeed, this is a satisfactory reductionist account of this domain, and in fact not lossy, up to the resolution specified by the model," as opposed to "this is predictive but known to be lossy"? What reason do you have to believe the ontology you chose accounts for all of the behavior observed in the actual domain?

I don't think this point is marginal. I see this mistake being made constantly, even by fairly sophisticated people, namely of too readily taking an ontology or a broader model as comprehensive before that has been demonstrated. Probably, considering that even physics, the reductionist domain par excellence, is not quite "done", we should be pretty skeptical of any claims to have completed a reductionist account of any other empirical domain!

Two examples which I'm going to harp on throughout this book:

- Everything that intersects with the study of the mind. Again, often researchers at the actual leading edge are much more parsimonious in their claims, but the point is well enough made anyhow.
  - It seems like quite a lot of people take the computational theory of consciousness as a fait accompli, even still to this day in the absence of a materialist account of consciousness! (Note that I mostly expect CTC to win.)
- I often see utilitarian and decision-theoretic models being deployed as a fundamental theories of value. It seems to me that decision theorists have largely skipped the work of justifying that you *can even* reduce human value to a consistent utility function. Worse, in many cases they recognize that human value is not consistent (in this narrow sense) and go on merrily with their philosophizing.

In these cases I want very strongly to say "hold up there, who ever said that was comprehensive?" To be clear, however: sometimes one has a research program, and indeed it's very useful for a person or group to pursue some set of assumptions while they still remain provisional. That's not what I'm objecting to.

Often I think one should be asking something like, "is this the kind of domain where we're going to get a nice comprehensive account, or is our intuitive ontology made out of so many heuristics and kludges, and if you try to hammer it into something flat and consistent you'll just be throwing out important functional aspects of what's normally regarded as the domain?"

I'll also say, even in domains in which I expect a comprehensive model is possible, it will be at a much higher resolution than that at which people are yet trying. Think of the riotous complexity of infectious diseases, which (my impression is) we having fairly straightforward models of at this point. People have of course been trying to understand disease for the entire history of our species, and we were working with *almost complete nonsense* for a large category of diseases until less than 200 years ago, and the picture continues to get messier and more complex as we learn more.

## Footnotes

1. David Chapman. [Meaningness](#) ←

2. David Chapman. *Metarationality* ↩

3. Hasok Chang, 2004. *Inventing Temperature*, pp. 8-11 ↩



# Good, Right, & Wrong

## ⓘ This perspective is not quite meant absolutely

There's no way to jump into meta-metaphysics without assuming some kind of metaphysics! I [argue later](#) that we mostly don't have frameworks for resolving metaphysical disagreements.

Even for those whose metaphysical beliefs don't agree with the assumptions I make here, I still will say that you can notice what distinctions these arguments make, which I'll be coming back to throughout the book.

I also think these arguments aren't quite airtight and are more to gesture in a direction. Still I think one broadly speaking can't really "get away with" making arguments, under most systems of metaphysics, that are fallacious under this model.

For moral realists in particular, I dig into a weaker form of this argument near the end of this section.

There are a variety of contexts in which people invoke concepts like "true" or "right" or "good," whether implicitly or explicitly. Quite a bit of reasoning about this is straightforwardly *incorrect* because conflates different kinds of *rightness*.

**My stance here is broadly spacious, personal, and flexible. I don't think you ought to feel or perceive what I do.**

A main thread throughout this book is that what's good and right and beautiful is going to be highly spacious, and allow for a great deal of flexibility in personal temperament, sensitivity, and aspiration. *On the other hand*, there are many claims that are made about what's right which are from what I can tell, metaphysically or factually *wrong*, and I want to criticize them very clearly.

## A sketch of an ontology of rightness

I'll say again, this is a sketch, it gestures in a direction, but it's not quite an airtight argument. In addition, this isn't trying to be an exhaustive ontology of "rightness", but is trying to point out one important dimension on which certain kinds of claims can vary.

### Mathematical and logical truth

In some sense we can start here with mathematical truth. Theorems are true in a solipsistic universe of meaning, they have no *certain* referent in the physical world, they are "true" only within their axioms and laws of inference. This doesn't make them weak, it in fact makes mathematical truths extremely robust.

Mathematical statements are true insofar as they are *logically necessary*, that is, that they *cannot but* be true, as a consequence of their assumptions. Mathematical truths are however often conflated with empirical claims, facts, etc., and I want to distinguish them very clearly.

## Empirical

Empirical claims, including either variously "bare facts," models with some assigned credence, or predictions about future events, all of these are constructed with respect to a physical (or at least external, etc.) world. These always have the shape, "one, or some, external objects, have a property or relationship, which (at least in principle) can be observed." Empirical claims need to be *demonstrated* to abide by some logical or mathematical representation.

Really there are quite a few subtypes here, which I won't go into. In general, though, I would say that empirical claims are "compound" with respect to mathematical propositions external or physical phenomena. When we talk about things having properties, sets of objects, shapes, numbers, and the like, we're engaging in fundamentally mathematical thinking.

Don't be mistaken, empirical claims are completely ordinary. Even though I'm expressing it in technical language, this is all humdrum and everyday: "Joe was upset today", "I broke my foot", "Lettuce isn't very filling," "I think Russia won't invade Ukraine," and so on, these are *empirical* claims.

## Evaluative

*Evaluative* claims will either look like "this is (dis)preferred" or "given a specific set of criteria, this is preferred," depending on whether they are stated in absolute or qualified terms. Evaluative claims are built on top of empirical ones, adding an *evaluative function*.

Some examples:

- "You played the song wrong," or especially "I played a wrong note"
- "That was a great move" (in a chess game)
- "That business decision was a mistake"
- "This painting is gorgeous"
- "Avocados are gross"
- "AI should not xyz"

Let's say, within evaluative claims, there's going to be quite a few subtypes. Among these, at least, we're going to want:

- Conformance, whether strict or relative, to a specification
- Optimality or fitness under a utility function or a loss function
  - A utility function or a loss function, for this purpose, is just a function which maps an input, whether a physical object, an idea, a strategy, etc., to a number. We can then rank inputs by the value of their output, eg. a machine which produces more widgets per hour, or which uses less fuel per mile, etc.
  - One special subtype would also be optimality or fitness of a *strategy*.

The concept of an evaluative function is very general; basically, we need to be able to distinguish *the object under consideration* from *the criteria used to evaluate it*. Mathematically, we can use different kinds of

functions with different properties for the sake of comparison and choice.

In practice, most people don't think of evaluative reasoning mathematically, and might balk at being asked to actually assign a number to some object or another. "Yes, this soup was delicious, but no it's not a matter of numbers!" etc. Nonetheless we can, broadly speaking, say that evaluative judgements conform to some mathematical structure. For example, if between objects A and B, either one is "better," or they are the same, then we have what's called a total ordering over evaluations.

(It's important to note that many evaluative judgements define what's called a partial ordering, rather than a total ordering. Eg. if asked which was better, the paintings of Michelangelo or the music of Stevie Wonder, most people would say "apples and oranges!" etc. Some evaluative functions are binary and don't define an ordering, and there are probably more nuanced cases as well.)

Evaluative claims can be said absolutely, or qualified. An absolute evaluative statement will usually not refer explicitly to an evaluative function. Evaluative claims can on the other hand be qualified in such a way that they become empirical claims which contain reference to an evaluative function.

I would claim that it's basically *meaningless* to refer to something as "good" or "bad" without a specified evaluative function. In practice, of course, eliding the evaluative function from an utterance like this is perfectly fine in a lot of contexts, and the point here is not to be the language police. *At least* I would say, we can notice when someone is making an evaluative claim, and notice that they are using an evaluative function, usually implicitly, and surface that distinction.

**One cannot make an evaluative statement, and then claim that it is true absolutely and *without reference to an evaluative function*.** If someone says, "this is the best car," we can ask, "good according to what?" Hearing an evaluative statement without an evaluative function should (I'll claim) be as strange as hearing a comparative statement without a point of reference: "this car is 50% better!"—better than what?

There is a sense in which one can say "indeed, a kilo weighs more than a pound," and assume that this ought to be basically utterly compelling to any sane person, but one *cannot* in general say "this thing is bad" in quite that same sense.

Evaluative claims are often understood to be *matters of taste*, that is, even if the evaluative function is left out, there's an understanding that people will vary and feel quite differently about food, music, etc. Sometimes this is a polite mask over quite a lot of judgement, but often no one has anything at stake, one simply doesn't like licorice, or noir movies, and has such preferences without thinking those things are wrong or bad.

On the other hand, the speech act of making an evaluative claim also often serves to, socially, make those who don't use the speaker's evaluative function dumb, low status, or evil, or to do the opposite for those who do. For example:

- "Rap music is garbage"
- "People who voted for [candidate] are idiots"
- "Homosexual sex is an affront on God"
- "Modern art is pretentious bullshit"

Other kinds of evaluative claims are implicitly claims about strategic choices, broadly speaking. These make claims at the same time about what goals one *should* have (again, what evaluative function one should use)

and that a particular strategy rates poorly under their evaluative function. These usually take the form of prescriptions, implicit or explicit. For example:

- "Carbs are bad for you"
- "You should finish college"
- "Regular exercise is what your body needs"
- "It's important to learn a second language"

Often these are less problematic; we can make some plausible assumptions about what values people *do* have, and say something to the effect, implicitly, of "assuming, as is true for almost everyone, that you want to be energetic, mobile, and feeling good on a day to day basis and into old age, exercise is indeed vital." Nonetheless prescriptions like these often sort of "cheat" by presenting a strategy as an unalloyed good, and leave out a specific person's desires, and especially their constraints. Once again, it is simply *incorrect* to say "this is a choice that should be made irrespective of all goals."

That said, notions like "goals" and "values" can often be myopic or shallow. These can change, and changes are often regarded as "truer" or "better." I think there's a lot of real substance to this, and that there's big questions here. I talk about in depth in the [last major section of this book](#).

### **More detail on the metaphysics of empirical and evaluative claims**

There's a useful sense in which "empirical" claims are contiguous with evaluative statements.

I would say here that one's ontology is convolved with one's goals or values. The notion here is that there is no sense in which an ontology can be defined, or maybe "would be defined," absent some values. More confidently, in at least *most* cases an ontology and some goals are defined together at once, so whether something "is" a member of a category has mostly to do with what you're using that category for.

In this sense, still, there's in most contexts at least a meaningful difference of degree between most empirical and evaluative claims, where the axis in which they vary is something like "how commonly shared the ontology is"—which even for empirical claims obviously is seldom 100% among humans, and sometimes quite a bit less. Again, in addition, see [no universally compelling arguments](#).

Modern psychiatric diagnosis is of course the standard counterexample here of what's presented as though it were ontologically straightforward, like *most* medical diagnoses, but is highly convolved with a specific set of goals, namely treatment, primarily pharmacological, and usually defined in terms of "difficulty coping" or similar. To be clear, I actually think the DSM (the standard diagnostic manual in the US) is doing a pretty good job here, given the various constraints and pressures. I think the main problem with psychiatric diagnoses are people who think they're something they aren't.

(IIRC in Quine's work there's another expression of a very similar distinction, namely of observations being "theory-laden.")

In this sense, my point that evaluative claims without an evaluative function are meaningless (really, a type error) also applies to empirical claims; we would have to say something like "under this epistemology" to be more absolute in our empirical claims. Indeed, I've sort of [already said as much](#), that the epistemology we use affects what we think is real and true.

Still, there's obviously a pretty large difference of degree between empirical and evaluative claims. Maybe I can't say "speaking an evaluative claim as though it were empirical is a type error" but nonetheless, I'm going to put my foot down and say that conflating the two is confused and often harmful. I think, given a shared frame, that it's broadly fine to make evaluative claims authoritatively, but that in domains where the evaluative function is contested, speaking as though the evaluative function is definite, or shared among all speakers, or actually just-as-real as empirical truths, is a kind of metaphysical and memetic harm.

[I also discuss](#) how these harms are perpetuated.

Finally, I'll talk [later in more detail](#) about the sense of empirical truth that I use and think is broadly helpful and appropriate, as well as [limitations of that model](#) (And, for what it's worth, [I end up endorsing](#) a kind of weak epistemic and ontological anti-realism.)

## Moral

Moral claims are a confusing point, and I'm not sure where I stand on them. In a meaning-reductionist view, moral rightness is purely just an evaluative claim; I'm not fully on this side. Part of the emotional, imaginal, and memetic power of moral claims comes from them being made absolutely. Taking morality purely as object often makes it thin and tawdry. Nonetheless I believe that morality is a property of minds, and not of the mind-external universe, despite the [claims of many religious traditions](#), and even some secular moral realists. (Relatedly, see [no universally compelling arguments](#).)

Briefly: moral realism is from my perspective usually something like [eternalist](#) window-dressing on real, legitimate, and vital moral sensibility.

In any case, moral truths strike one as more important and less flexible than optimization, efficacy, or beauty—and certainly not just a matter of personal opinion or taste. Murder is wrong! Rape is wrong! Torture is wrong!

(I'm pretty sure I really mean that. I've never witnessed these things myself but I'm pretty confident that if I saw them in front of me, I would be very disturbed, regardless of whether I would have the courage to do anything.)

But, there's moral contention! Many people in the world have believed, and still believe, that homosexual sex is evil, while many people in the liberal west think it's just fine. And so on for countless other examples.

For now, I'll say, I still think that strictly speaking these are evaluative claims, and are constrained in their "objectivity" (read, maybe: "pure inescapability") in the same sorts of ways. I also think that this doesn't actually need to weaken their emotional weight, at least in general. (Often it does in particular, for moral claims that have been taken as more real than they ought to have.)

I'll say as well, there's moral change if not moral progress. There's at least maturity, attunement, and alignment with one's deeper truth. The obstructions to our virtue and goodness can be cleared, and we can have truer apprehension of what is good and right. I think some moral anti-realists refuse this—I don't, and I think moral growth and moral insight are real and important, I just don't place them in an eternalism.

I'll discuss [at much greater length](#) how metaphysics relates to and interacts with various kinds of moral claims. Among other things I talk about why believing that morality exists in minds doesn't necessarily imply or require moral nihilism.

## What should moral realists take from this?

For died-in-the-wool moral realists, there's a few points I'll stress:

1. It's not clear *which* domains of evaluative claims are under the purview of your ultimate moral reality.

Obviously we've come and gone on this point culturally, eg. regarding apostasy, lots of details of sexual ethics in particular, capital punishment, etc. Maybe you can say "indeed, there's some moral uncertainty, but certainly murder is wrong." I'll grant you that, but this brings us to the next point:

2. Even if we assume moral realism, you *still* have to demonstrate that any particular moral claim is true.

Furthermore, the means by which we do moral investigation and discernment are mostly the same anyhow, namely introspection and debate. I have the same apparatus for ascertaining moral truth whether moral realism is true or not, my mind and my heart. So then, we're largely back where we started: there's moral uncertainty, and the metaphysics is ambiguous and difficult to determine, except insofar as you believe you have perfect gnosis of moral truth.

(Some may or may not want to count prayer as separate from introspection. Separately, even if your moral epistemology rests on a textual tradition, there's still an enormous amount of moral uncertainty due to hermeneutic problems, in which case the above still applies. If you think your epistemology and tradition of biblical interpretation are unassailable, then I suppose I can't convince you of much.)

3. Everything not covered by your moral theory is subject to the limitations of evaluative claims in principle that I've laid out.

You can't *really* trick someone into not wanting something that they do, at most you can either bully them into denying their own feelings, even to themselves, or you can cause them to notice what's good or bad about something, using their own discernment.

## Opinions, matters of taste, and ways of seeing

One of the important questions in judging some claim will often be "is this domain, or this type of evidence or argument, applicable to the kind or degree of rightness being claimed?"

For example, some questions are really a [matter of taste](#). Or, even if there is an "objective" structure which distinguishes one thing from another, like that something is more complex, precisely tuned, or difficult to achieve, *still* you might just actually enjoy the simpler thing more, and your preferences can't meaningfully be said to be "wrong."

I would say that a nearby category would be [ways of seeing](#). I wouldn't call these a *matter of taste*, but it seems like, various *consistent* ways of seeing, while powerful and deeply felt, can't really be distinguished as "wrong" or "right," but more like "helpful," or "appropriate," or "clarifying."

This obviously confuses matters even more about moral claims. I still have intuitions along the lines of "no, murder really is wrong, that's not just my feelings"—but this is a real question for me. However, there are cultures and legal systems in which eg. honor killings are acceptable and legal. I'm mostly fine with saying

that they're wrong too, but it's often not clear what really counts in here, and I still don't think we are [able to determine or adjudicate](#) this meaningfully.

# (Mis)use of Language and Rhetoric, re Rightness

---

There's a common kind of rhetorical trick, in which someone sneaks an [evaluative claim](#) in, like "a wolf in sheep's clothes," as though it were certain and absolute. They then carry on to convince those listening of something which they couldn't argue for if it were clear that this claim isn't *true* in the same way, or with the same force as *stabler* kinds of truth.

Colloquially there are quite a few different words that are used to refer to different kinds of evaluative statements. From a technical perspective, a musical performance might be described as "wrong" or "poor," variously on the basis of strict adherence to the specification (the written piece of music) and of the performance's expressiveness, beauty, etc. Optimality and fitness often take quite a few adjectives, eg. "beautiful," "delicious", or for that matter "sexy," but also of course just "good" or "best." Strategic fitness in particular is often described as "right" (as in, "the right choice"), "smart" (same) or "effective."

"True," even when used consistently, means something fairly different for mathematical claims as it does for empirical ones. We also commonly use the words "right" and "wrong," though never "false" or "incorrect," to describe evil, sins, or moral failings.

Again, without becoming needlessly pedantic, we can *notice* that there are differences between different kinds of language for different kinds of rightness, and that words have a prototypical meaning. **I think that one can "cheat" rhetorically by using one set of language which brings in intuitions about the strength and stability of one kind of rightness to bolster a different kind of claim.** Furthermore, I'd say this isn't just about manipulative rhetoric, often the language a broad culture uses to describe one kind of *rightness* implicitly conflates it with another, eg. "this is the truth" (of a spiritual teaching).

Evaluative claims are often spoken with the same tone and language as empirical or logical claims, and as though they had the same "inescapability." There are contexts in which you can reasonably say "these are pretty much just the facts of the matter," which serve to say, ostensibly, "you ought to be compelled to believe this, due to the kind and strength of evidence for it." I believe it is usually inappropriate to present evaluative claims as though they have this same kind of compelling force.

I think people are sensitive to this kind of ruse only *to some extent*. I see quite a lot rhetoric, broadly in the context of morality and values, which makes this move and meets with no open objection.

I'll make here a moral claim: in the same way that it's wrong to deceive someone about some *facts*, whether by straightforward falsehood or by manipulative rhetoric, it's also wrong to cause them to believe evaluative claims using the sorts of rhetorical abuses I've described. Insofar as you respect the discernment, autonomy, and agency of the people you're speaking to, you should make evaluative claims transparently, and let people decide on their own two feet.



Now, there's nuance, and there are exceptions. The more that a frame is shared within some community, or the more that there's effectively consensus about what evaluative functions to use, the less it matters if we leave the evaluative functions out when making evaluative claims. It would be ridiculous to call it "rhetorically abusive" for a religious order, within its own community, to simply have shared assumptions about what is good and right. Nonetheless I think there's something pernicious about the "unknown knowns" of ideology and metaphysics, where the entire evaluative frame remains tacit, even if it's not contentious.

And then, what should we do when the evaluative frame *is* contested? Descriptively, when talking to others that one trusts and respects, people will generally surface the evaluative function as contentious. "In my opinion," "I've been told that," "the experts say x, but I'm not sure," and so on. I don't have a smack-down argument, but my sense here is that insofar as we want to cultivate a [productive, healthy discourse](#), this kind of nuance and openness with respect to ambiguity and contention is broadly what we should be striving for.

(I'll also say, there's some rabbit hole here about socializing children, for which I just don't have an answer. One way or another, children will absorb moral intuitions, some of this is embodied and demonstrated, but a lot of it is also taught explicitly, and it seems pretty much impossible to try to teach precepts to children in a way that's nuanced with respect to a broader cultural discourse.)

One last piece: often people conflate, implicitly or explicitly, different kinds of evaluative claims. In particular often claims about what strategies are effective end up transmitted culturally in terms of morality, obligation, and sin. I'm thinking here about many people's relationship to exercise, diet, parenting, "financial health," etc. This will also be an important distinction to keep in mind when we talk about [spiritual practice](#).

## Taking a specific example: confusions regarding aesthetics

*First: the point here is to say, "here's a nuanced case of misunderstanding the role and strength of certain kinds of evaluative claims." When we get to talking about [spiritual truth](#) later, I'll return to this question, of what kinds of truth and correctness apply when talking about morality, goodness, spiritual teachings, etc.*

In music we have the concept of a "wrong note." A piece of music, as written or as transmitted in an oral tradition, defines a particular sequence of notes for certain lengths of time, and of course people constantly make mistakes and simply play a different note than what is expected. This sense of "wrong" is fine enough, and indeed in many contexts in music we just want to play the piece as written, or at least as written and interpreted within the bounds of the tradition. In fact in those contexts everyone will basically agree as to what "wrong" means, and what counts as a "wrong" note.

However, this evaluative function is only suitable in some contexts and not others. In jazz, many of the rules are relaxed, not into chaos but into a denser and richer structure. You will literally find jazz musicians saying that there's no such thing as a wrong note.<sup>1 2</sup> In jazz improvization, if you play a wrong note, you can just about always recover it into a "right" note, with enough skill. Separately, much of the history of Western classical music, from Bach until Schoenberg in the early 20th century was a progressive elaboration and *subversion* of established structures. Most of the classical composers you know made their name by doing things that were "wrong" at the time, in tasteful and beautiful ways. This tendency can get arbitrarily wacky, see again Schoenberg's atonal work, free jazz, microtonal music, and even harsh noise<sup>3</sup>.

Here's a marvelous quote on the subject from Plato<sup>4</sup>:

Thus their folly led them unintentionally to slander their profession by the assumption that in music there is no such thing as a right and a wrong, the right standard of judgment being the pleasure given to the hearer, be he high or low. By compositions of such a kind and discourse to the same effect, they naturally inspired the multitude with contempt of musical law, and a conceit of their own competence as judges.

It seems obvious to me that there's something quite confused about this take. Maybe we can argue about good or bad taste, but "right" and "wrong" (note that another translation<sup>5</sup> gives "as a thing without any standard of correctness") don't even *make sense* to me to apply to music. I would say that the goodness or rightness of music is a "matter of taste," that there *isn't going to be* something like a "right" answer, in the way that we find with mathematical or empirical claims.

This is a nice example to dig into because, I expect, almost everyone reading this will agree that Plato is doing something confused here. This is not just to say that he's mistaken about some empirical fact, but that his metaphysics is deeply confused, and perhaps *wrong*.

(Even if you believe that some people just have atrocious taste in music, it's still hard to argue that music can be "right" or "wrong" in a very strong sense here. If you disagree, imagine, as an example, if someone said that burritos were "wrong." Fine as hyperbole, but if this is said fully in earnest, clearly it's gotten something *wrong* metaphysically.)

More generally, we can ask, "in this domain, what kind of rightness do we observe?" In particular if we're concerned with evaluative claims, again, "what kind?" Is this a matter of taste? Is this about what choices are strategic, given some set of goals, but where the goals are "up to" each person? Do we believe that something is good or bad in a more stable sense, like morality (if we believe in morality as stable at all)?

To be clear, even in the weakest sense, to say that something is a *matter of taste* is not to say that it *doesn't matter*. A piece of music might be utterly lovely or even gobsmacking, even if I'm confident in saying that I'm no more "right" in finding it so than the people who don't, or that anyone can be "compelled by the evidence" to find it lovely, the way that they ought to be *compelled* by a mathematical argument.

## Footnotes

1. [Jacob Collier](#) ↩
2. [Victor Wooten](#) ↩
3. [OPN](#) ↩
4. Plato, *Laws* [700e](#) , trans. Hamilton & Cairns 1961 ↩
5. Plato, *Laws*. [700e](#) , trans. Bury 1967 ↩

# "Real"

---

Quite a lot of jousting goes back and forth about what is "real." I will claim that the determination of some phenomenon as "real" is often a stand-in for questions like "does x matter," or "should we pay x any mind." Furthermore, I'll claim that the fact of the matter is nuanced enough that often whether something is "real" doesn't *really* help us answer these questions.

## ***How an algorithm feels from inside and disguised queries***

I'll bring in two concepts from Eliezer: how an algorithm feels from inside<sup>1</sup>, and disguised queries<sup>2</sup>.

Most concepts are [nebulous](#), but still track real structure in the world. Human intuitions about membership in a category often construct membership as an *extra fact*, besides all of the facts about an object's properties.

In the proverbial "if a tree falls in the forest, but no one is around to hear it," there are two properties which are normally tightly correlated, namely an event causing acoustic vibration, and it causing aural perception/experience, that have been decorrelated. This question is sort of about asking, "but did it really make a sound?" as though there were a question left to ask, besides the two component phenomena. Yes vibrations, no perception, simple as. To be clear, I don't think anyone is actually vexed by this question, but it's helpful to lay it out very precisely, because indeed I see this exact mistake being made with respect to this class of problem in general.

Another example: are viruses alive, or for that matter [prions](#)? Viruses do not have a metabolism and can't reproduce outside of a host, but do contain genetic material, and do evolve. Prions are much simpler, they are misfolded forms of ordinary proteins, they don't contain genetic material, and they only reproduce by causing normal proteins to misfold. Again, we might ask, "but are viruses really alive?", as though there were a meaningful answer to this question, beyond the facts of which properties they have in common with most kinds of life, and which they don't.

We can construct a category in such a way that marginal cases are strictly speaking not members, but we haven't really *learned* anything by this. There is a large diversity of observed biological systems, and gerrymandering "life" to exclude some of them hasn't told us anything new about the ones we excluded, nor would it to arbitrarily choose to include them.

In Eliezer's explanation, his claim is that we might imagine choosing one algorithm (or really, data structure) or another to represent some concept, and that the intuitions that humans have, in which this "but is it really?" question arises, would happen in some kinds of representations and not others. As well, these kinds of ontological intuitions are constructed in his view out of cognitive algorithms, and that the intuitions are "what the algorithm feels like from inside of it."

Eliezer's claim in *Disguised Queries* is that questions like "is x in category y" are often stand-ins for "is x dangerous," or "is x reasonable," or "is x important." A category is presumably predictive of some other properties, besides those that define the category. Often people will haggle over whether x should be

included in a category, *in order* to be able to determine that we should infer some other property about x. For example, we could (well, at least in principle) imagine someone trying to argue that Gandhi was a criminal—which is absolutely true, strictly speaking! Of course, our imagined speaker here is really trying to argue that Gandhi was dangerous, or immoral, or whatever. This is a silly example, but this is *absolutely* a move that people make with respect to categories like racism, violence, and terrorism. These kinds of questions are called "disguised queries," since the real question which is being debated is disguised by a question about where to draw a boundary.

Let's assume, though we don't necessarily need to assume this, that the reality described by physics is stably "real" in a straightforward sense. I'll make a few claims:

- There's a fairly large diversity of phenomena which we call "real" (or at least which we *regard* as real) that are **ontologically** very different from physical reality.
- Many of the details of physics are quite counterintuitive, and the deepest questions about what's "real" are still contentious in physics.
  - Physical solidity (as in, of wood or stone or bone) is quite a bit subtler and stranger than ordinary human-scale intuitions about "solid matter", though my understanding is that the popsci schtick that matter is "mostly empty" is somewhere between oversimplified and false.
  - Famously, the cosmological interpretation of quantum mechanics is *still* contentious even in academic physics. Furthermore, both interpretations currently in the ring get openly metaphysical, and in either case are quite contrary to day-to-day intuition.
- I don't believe that there's an account of what's "real" which includes far-flung notions like romantic love, money, status, mathematics, or music, that doesn't also end up including a pretty wide swath of meditative and spiritual experiences.
  - Perhaps, we can hypothesize various grades, or at least a diversity of disjoint types of *reality*, which is broadly the stance I like, but then we end up with "real" as something either much weaker or much narrower than people usually want it to be. (Also, gosh, wtf, *multiple realities*? Is that really how far we've stooped?)
    - In any case, this doesn't really help, because there still doesn't end up being space to argue that all this spiritual stuff is just "not real."

I want to be very clear here: I am not claiming that nothing is real, or that reality is completely socially constructed, or that everything which we call "real" has the *same* ontological status. I am claiming that "real" is both nuanced enough and diverse enough ontologically that it *must* admit spiritual experiences, one way or another.

### **And what about this assumption that physics is real?**

Look, if we're going to throw that out then you're just *really* screwed.

More seriously, if we don't assume a basis of reality like "physics," then whatever sense of reality we're working with is yet weaker, and so my broad take is even stronger.

I suppose this obviously leaves out gnostics, or various "mind-only" or "mind-first" metaphysicses. I'm not really writing for those people, but in any case I think they're already going to agree with some claim like "reality is more nuanced and ontologically fussy than modern secular metaphysics would like."

Let's take some specific problematic cases:

- **Math:** There have been a variety of theories with respect to the ontological status of mathematics, over millenia. Math seems both more inescapable than physics, but also completely intangible. We can *imagine* (and, it's now often hypothesized) worlds with different laws of physics, but we largely *can't* imagine "worlds" with different arithmetic, or calculus, or topology.
- **Music:** Is music real?—sorry, wait what? Since when is music even the kind of thing that can be real?—well, we care about music, we devote quite a lot of economic productivity to it, we argue about what music is right, and some people dedicate their lives to the practice of it. For most people it's often much more salient and powerful than math or history or certainly theories of physics. On the other hand, it's not quite like most things we would ordinarily call "real." It's just sounds, often without words, in some pattern, how could that matter? How could that possible mean anything?
- **Love:** Yes yes, "true love" and all that; but in any case, once again, is love real? What would it mean for it to be "real"? Sometimes people feel bamboozled or manipulated by love, and there's certainly deception and fraud presented as though it were love. On the other hand, love, whether romantic, familial, love of one's country, etc., seems like one of the most important forces in world. People will give their lives for it, they'll work themselves to the bone for it, etc. etc. etc. Love is also richly imaginal: tied into fantasy, imagery, and elaboration that are both *more* salient, powerful, and motivating than what's actually in front of our faces, but also never quite present "in the here and now."
- **Money:** The value of money is certainly "**rational**" in a clear sense: macroeconomic shenanigans aside, you can generally expect that you will find a counterparty willing to accept your dollars for most kinds of transactions. Still, money isn't quite just "made out of physics," or at least not *just* physics. There's a large, only-somewhat-centralized cooperation in the construction of a coordination mechanism, which relies on expectation across the network that the rest of the network will continue to exist, and continue to behave largely the same. Of course, money is quite real! Money is incredibly salient and powerful, almost anywhere in the world. So we have both a very high degree of indirection, and something to the effect of "social construction," which is meanwhile quite "real," in the sense of "important, and worth worrying about."

So, again, if a tree falls in the forest, but there's no one around to hear it, is it real?—sorry, again, there's a variety of properties, some in common with other things which we take as real, some not, and no extra fact about whether it's *actually* real.

## Ontological and spiritual commitments

An [ontology](#), among other things, serves as the framework through which we understand what exists, and what *can matter at all*. Descriptively, most people have commitments to particular ontologies: moving from one ontology to another often inspires a fear that what one now understands to be real and will become meaningless, wrong, or harmful, and vice versa. This is often correct! One metaphysics will often, in fact, [reify](#) or valorize notions which are fake or irrelevant in another.

These kinds of ontological commitments might be made by way of organized religions, ideologies and philosophies, or may be cultural and implicit, idiosyncratic, and even completely unspoken. Some examples would be things like reductionist materialism, Karma<sup>3</sup>, Heaven and Hell, souls and spirits, and even persistent personal identity.

To be clear: this list is not to say "these are utter nonsense, and everything that flows from them is wrong." Rather, mostly I want to inspire some sort of ontological flexibility and a skepticism of over-reified systems of

metaphysics. (Still, as an empirical claim about cosmology, I'm pretty confident that Heaven and Hell are not real.)

Most of the point here is to say, a. that the ontologies that people use are largely *unjustified* (though in the [next section](#) I'll get to why this is so hard), and b. that in general, these domains are nebulous enough that there isn't going to be a singular, stable ontology on which to build all of our theories of what matters or is worthwhile<sup>4</sup>.

The naive view which says "physics!" or "science!" [in response](#) to questions of ontological nebulosity simply can't make these problems go away. But, still, there are things which aren't real! Certainly, there are lies, and there are dreams, and there are hopes that do not come true.

Often, scientism and atheism maintain a commitment that entertaining many [ways of seeing](#) is "wrong", since these are "not real." Except, again, a. what's real, but more importantly, b. how have you determined that the collection of ways of seeing that you use are correct, or the only ones that should be used?

Now, my real question is: what should we make of both claims and experiences of prayer, meditation, psychedelics, mysticism, etc.? I think these are real questions, I mostly don't have a clear answer. One of the main claims of this book is something like, "actually, I still don't have a clear answer, but it's definitely not any of these."

Note that I'm not claiming that eg. those who report talking to God are definitely talking to a persistent physical being etc. etc. Mostly, I think these experiences often have very substantial existential value, and can't be excluded. However, determining which are baloney and which are actually worthwhile is much harder. [I discuss this more later.](#)

(Though, I do have a very small probability on some spiritual reports having a straightforwardly literal ontological/cosmological interpretation. Again, [I discuss this more later.](#))

I'll note, that if I say, "well, there's something interesting going on metaphysically with respect to spiritual stuff," this isn't yet to say that you should regard it as important or valuable, but just to say that it can't be flatly determined to be not real, just by writing it out of your map.

I'll discuss this all in much more detail in the section on [spirituality and spiritual practice](#).

## "Real" and "True"

One last bit here: I'll claim that "true" is nuanced in basically the same sorts of ways as "real" above. What's true? Just mathematical theorems? What about very high confidence predictions, from scientific theories? What about just *high* confidence predictions? And, which theories, or is it mostly about their being socially acceptable, and endorsed by the current priesthood? What methods are even acceptable for determining what's true? What does it mean, if anything, for introspective or spiritual claims to be true?

## Footnotes

1. Eliezer Yudkowsky, 2008. [How An Algorithm Feels From Inside](#) ↩
2. Eliezer Yudkowsky, 2008. [Disguised Queries](#) ↩

3. This is usually left out of bowdlerized western Buddhisms, but Buddhist Karma traditionally amounts to something to the effect that the goodness of actions is ontologically/cosmologically basic. I think it's a bit hard for us to appreciate both how weird and how intense such a perspective would be, if fully taken on. ↩

4. [Chat, is this real?](#) ↩

# Metaphysical Disagreement and Determination

---

There's quite a lot of disagreement about metaphysics in general, if not explicitly by that name, and about subjects downstream of metaphysics. I've perhaps *never* seen productive discussion about these topics. I have some ideas about why.

Some claims in this section:

1. Metaphysics matters.
2. It matters if you're wrong about metaphysics.
3. We do not have frameworks for doing metaphysical discernment, either personally or communally.
4. This should give us pause.

## Metaphysics matters

I've already said this in the [introduction to this chapter](#), but I'll elaborate on it some here.

Large classes of beliefs are downstream of metaphysical ones. These beliefs might be professed, or carefully reflected upon, or just simple and grounded and so intuitive that they're completely invisible. This really matters: **these beliefs determine our seeing, feeling, and acting in the world.**

Metaphysics is also not optional. There is no neutral perspective which is totally agnostic of metaphysics. One has a metaphysics, or at least some vague collection of metaphysical intuitions, whether one would like to or not. In fact, if you didn't have a metaphysics, you would be unable to learn, to infer, or to want.

Even some of the wackier pie-in-the-sky consequences of various metaphysical systems are relevant to your ordinary life and experience: most likely you want to keep living tomorrow, and you wouldn't want an alien to take over your body, or for your body to keep operating but for your conscious experience to end. You also probably expect that your life will end, likely after less than a century, but even if you expect human life extension to succeed, then you still probably expect that your consciousness will end at least due to heat death of the universe. Both transhumanism/futurism and Buddhism have claims that affect all of these, and one way or another you have to grapple with that, even if just to say "eh, I'm not gonna take this seriously, don't bother me about it."

If personal identity is confused or not real, that matters enormously. If rebirth is true, that matters enormously. Conditional on cryonics and mind upload succeeding, those matter enormously for how you live your life today. As philosophical or empirical claims, you can dismiss these, but you can't claim that the metaphysics of minds, for example, is irrelevant or not real. (Personal identity is probably the strongest one here, since it matters on a moment-by-moment basis, not just some decades from now when you die.)



Metaphysics is also very mundane. Though we highlight metaphysics mostly when it gets wacky, it's also just part of the warp and weft of [ordinary experience](#). More saliently, also, metaphysics is especially fundamental (if subliminal) in the ways we understand the value and purpose both of our own lives and of all life, the universe, etc.

Lastly, lots of object-level debates often have upstream disagreements about metaphysics. Some obvious examples here would be abortion and trans rights. To a decent extent these debates can be understood as mostly being symbolic contentions between two tribes, *however* there I think are [real disagreements](#) about the metaphysics of life and birth, and of gender and sex. Both of these are inherently [nebulous](#), and each side has declared a determined, correct metaphysics here, while mostly not surfacing the metaphysical questions here, as such.

### **Weakly contra the Durkheimian hypothesis**

I'll discuss this [more later](#), but briefly: one hypothesis which we might have for stories about cosmology, value, truth, etc., is that they mostly serve as loci of social coordination. In this hypothesis, we might say that both the ostensible [epistemic](#) content (the specific claims about the world) and the axiological content of religious and mythological stories is sort of "beside the point," and instead these exist to provide a story about which a community can agree, and which can be used as a basis for people to understand their roles and responsibilities, for them to come to consensus, and to strengthen their social bonds. We might also want to say that there are lots of equally functional stories about what's good and right, at least at the level of high-level social outcomes.

I think this is only partially correct. I'd say, yes, to a large extent the epistemic content is arbitrary, and on some important dimensions you'll probably get similar outcomes for a wide range of metaphysics, or at least that most of the variance is not explained by the metaphysics. However, the metaphysics a society uses still has real consequences on outcomes we care about, and certainly on apparently psychological weighty subjects. We might imagine one culture which says that its warriors will be remembered for their glory and honor, or another which says that they will live forever in heaven. Perhaps they both have similar levels of social cohesion, and similar levels of bravery from their soldiers. But come on, obviously I personally should care if I actually expect to live forever in heaven as reward for my valor—and in any case I should be seriously questioning if it's good to incentivize war at all.

I'd also say that the extreme perspective in this direction is a kind of nihilism: human beliefs and moral intuitions are really just selected for reproductive success, and locally speaking are socially incentivized to align with some or another power structure, so why are they meaningful at all? What obligation do I have to any moral expectations? Again I will discuss this [more later](#), but briefly I'd say: you almost certainly don't actually believe this. You still don't want to die, you still want love and safety, and you'd still be horrified at seeing yourself and your family tortured.

On the other hand, some kind of "meaning maximalism" is probably also wrong. Sometimes, those particular inclined to consistency and piety will draw out all of the consequences of a metaphysics and take them deadly seriously. I'm thinking here of everything from Anarchists to Jihadists to Effective Altruists. I think this ends up biting the bullet on all kinds of mundane value and meaning that we probably shouldn't want to throw out.

One take here is that a metaphysical claims and stances are more heuristic than absolute. The point being, they can stand for a sort of considerations, without being "taken literally." For example: often we plainly talk

about something being "bad for your soul", without really worrying about the metaphysics, to correctly point at a worthwhile set of considerations for what might also be called "moral injury" etc. I like this take, but I still think this doesn't really solve the problem. We actually still have the same question, namely, how do we apply some big collection of lossy heuristics to themselves, to arrive at even just *better* or *worse* answers for pressing questions and tradoffs. This is still very difficult and confusing.

## What does it matter if you're wrong?

With some caveats, I think it matters quite a bit if you're wrong about metaphysics, [for various senses of "wrong."](#)

First, some ways in which it matters less:

My impression is that by default cultures broadly, and people individually, don't develop a very sophisticated metaphysics. In my thin reading of anthropology, and particularly the anthropology of religion, it seems like the criteria for admission into a folk metaphysics are something like, "is it resonant," "is it simple enough that the culture can transmit it," and "does it not grate too much against existing sensibilities." Shooting from the hip, it seems like almost all folk metaphysicse I've ever read about are unsophisticated in ways that generally rhyme with one another.<sup>1</sup>

Folk metaphysicse are also often blatantly inconsistent, or they consist of some funny hodgepodge of systems which explicitly contradict each other, eg. many local shamanic systems in Latin Americae which overlap culturally with Catholicism, or the same more marginally with Islam. Even many attempts at careful, systematic metaphysicse often contain straightforward logical errors<sup>2</sup>.

And yet, the world goes on, people go about their lives, and society remains moderately functional—with whatever problems there are smelling more like "politics" or "economics" or "violence" than metaphysics.

We might also say, "look, there's lots of ways of coordinating society, every society will think itself *good*, and often *the best*, why quibble over their stories about ultimate value?" Or perhaps, "So, lots of people believe funny things, they pray to funny gods, but they're mostly having a good time, so what does it matter?" More strongly, we might also say, "even when people have been killed ostensibly due to religious differences, probably this has more to do human tribes warring than it does anyone's professed metaphysics."

I think these are all right, and nonetheless I think you still can't get away from metaphysics mattering. As long as you still care about death, as long as you still care about whether your experience is real and not a dream, or as long as you feel that torture or rape are abhorrent, you still care about metaphysics! (Or at least, you are about distinctions which are downstream of metaphysics, and so therefore in some sense you also "automatically" care about the upstream considerations.) Insofar as you believe in and care about things which would be repudiated by a different metaphysics, **you must care about your metaphysics, and you must care about getting your metaphysics "right,"** since indeed the contradictory metaphysics must be something to the effect of "wrong."

## Metaphysical discernment and determination

I am not aware of any frameworks or models which allow one to meaningfully investigate and come to determination about metaphysics. Even for a person working privately and only trying to figure out what's right, I'm aware of no process of metaphysical discernment which converges to consistent and correct

answers across different people, in the way that math or empirical disciplines can. (That is, without effectively just picking the answer from the outset.)

Looking at the diversity of systems of metaphysics around, I don't see any way for them to interface, to be falsified, or to find concrete points of disagreement which could cause a believer to change their mind.

This is partly because metaphysical beliefs are often unjustified, or worse *unjustifiable*, but also because the [domains of metaphysics](#) are inherently quite [nebulous](#).

I'll go through each of the domains and describe some places where what seems clearly defined becomes hazy and confused. I want to be clear that in each domain of metaphysics, I don't believe that metaphysical nebulosity lets you get away with whatever you like.

My main argument here is that if you look carefully, with an eye for nebulosity, the parts that are actually pinned down are frustratingly narrow, and that just about everything that relates to meaning and value is left flailing. Furthermore, all of these flailing pieces are *most of what really matters*, and we don't have frameworks for trying to inspect on intuitions and reflect, which would allow us to arrive at anything like *correct* or *satisfying* answers. I haven't seen a presentation of a particular metaphysics which doesn't end up saying much more than just "because I said so."

## Cosmology

This might seem the most concrete, empirical, falsifiable, and stable, but still I think there are problems and a frustrating amount of nuance.

There are indeed stars in the night sky, there's red-shift, there's the cosmic microwave background, and there are modern theories of cosmogenesis, the many-worlds interpretation, and even Tegmark's Mathematical Universe Hypothesis<sup>3</sup>. These range from flat, undeniable, and plain-to-the-naked-eye, to highly indirect but still well substantiated, to speculative and unfalsifiable.

Modern cosmology, for whatever uncertainty there is at the highest level, is presented as basically empirical and well substantiated, and broadly as having made traditional, pre-scientific cosmologies completely obsolete.

However, cosmology in almost all cultures (as I understand) has functioned less as an empirical theory of a natural phenomenon, as it has as a *story* to structure the place of humans in the universe, our responsibilities to each other, and our responsibilities to the symbols and sacred principles of a given society. I would call this the [imaginal](#) (yes, not *imaginary*) aspect of cosmology. Another way to say this is that the imaginal aspect of cosmology conveys [ways of seeing](#) which are often tightly integrated with the rest of a metaphysics.

To be clear, the traditional imaginal stuff is often stupid, insane, and harmful—I'm not being a naive hippy, saying "ah yes you should trust all traditional religious systems, they contain deep wisdom that the hyper-rational west has lost." I'm also not saying that it doesn't matter that the ostensible epistemic content is wrong. The distinction I'm making here is that, descriptively, cosmologies have *mostly* been about the imaginal aspect, rather than a concrete empirical theory. Indeed, the stars couldn't be interacted with, and until about the 16th century couldn't really be inspected more closely, besides perhaps tracking the planets carefully, or recording comets. Nor, for that matter, can we really "do" much even with modern scientific cosmology.

Even modern scientific cosmologies still do have imaginal consequences, though exactly what is not clear. Sometimes this gets turned into stories like "we are the universe understanding itself", and sometimes it's something like "we are puny and irrelevant and the universe is fundamentally uncaring." Another cluster of popular modern ones have the shape of a kind of futurist manifest destiny. In any case, the empirical claims will sort of "automatically" be used to construct an imaginal world.

Often they are much more than that: if you watch the TV show *Cosmos* (either the original or the new ones), astronomy and scientific cosmology is taught by way of the hagiographies of famous scientists. There's awe, and glory, and maybe even sacredness in it; the cosmos stands for the valorization (and justification) of scientism.

I'm not going to contest the actual empirical claims of scientific cosmology, though I do wish that popular discourse was more conscious of the uncertainty *in the discipline* of a lot of these models. The real question to ask here is: **what story should we tell about the history and fate of the universe, and our place in it?** As I'll be asking throughout this book: what ought we want to valorize? What ought we censure?

Note that this is not really about nebulosity with respect to the actual empirical claims as it is about meaning. Whatever certainty the relevant physics and astronomy support doesn't help us answer these questions. Furthermore, I don't think that ways of seeing can really be **said to be "right,"** rather than they might be helpful or miserable or lovely.

I would say also that the imaginal aspect is in many ways more important than the empirical aspect. Most people, insofar as they care about cosmology at all, do not care about specific models of physics but indeed mostly about how it makes them feel, and what story they feel that they're embedded in. Even the most staunch atheist can't really *escape* the imaginal consequences of their cosmology, as an important part of how their narrative world is structured, among other packages of meaning and narrative as well. I would say also that some traditional cosmologies often have a fair bit besides bathwater in them, and I think secularism has probably too hastily rejected all of them.

Now you might say, "wait, but you just said it matters if it's wrong, but now you're saying it sort of doesn't matter if the cosmology isn't really literally true?" I would say, yeah, it maybe doesn't matter if a cosmology is correct as in "literally true", and so it would perhaps less so be **"wrong"** as "harmful" or "bad." But then, perhaps moreso, it surely matters if our cosmology is wholesome, and helpful, rather than harmful or destructive.

## Ontology

As with the others, ontological nebulosity doesn't let you go full Humpty Dumpty<sup>4</sup>. Of course, in principle, you could define words however you like, and you could lump together this or that category, but it still wouldn't cause there to be some real structure in common with those. Often, we might start out with some set of intuitions about what kinds of things exist, and then by investigating them more closely, it becomes obvious that our prior intuitions were confused, and that there isn't much negotiation that's really possible about whether the earlier folk models were correct. Some obvious examples here might be the germ theory of disease, or the atomic structure of matter.

But, very much, most categories are really more like messy clusters, and our intuitions, whether they're "baked in" biologically (eg. lots of brain structures for identifying human faces) or learned, often balk or throw up their hands when faced with the complexity in the actual domain. This often matters quite a bit: when is a person alive, or dead? (Is this person really a woman?) Again I would stress here that an ontology is a

property of a model, or a view, rather than a property of the domain, perhaps (with some nuance) with the exception of whatever phenomena are really "at the bottom" of physics.

I'd mostly like to avoid openly political subjects, but of course I have to talk here about gender. A popular argument from the pro-trans side (leaving aside that 20 years ago public-facing pro-trans rhetoric was mostly pseudobiological and gender binary-ist, "born this way," "female brain," etc.) has been that even biological sex is pretty nebulous. To my understanding, as a matter of biological fact, even if you wanted to construct a totally gender-essentialist, binary-ist ontology, there are still quite a few edge cases in the real domain, namely basically "intersex" people. (You could, on the other hand, just declare intersex people to be abominations, or even non-intersex trans people, and the actual biology doesn't preclude that.)<sup>5</sup>

Broadly, often, people like to assume that the ontology used by "scientists" is comprehensive, authoritative, and exclusive. Again you'll find this in medicine, you'll find it hawked in both directions with respect to gender discourse, genomics, intelligence research, economics, etc. Often these "scientific" ontologies are some weird amalgamation of a few generations of scientific theory, vague folk theory, and dumbed down explanations for public consumption. My impression is that in many cases the real literature acknowledges quite a bit of nuance, despite what the general public things "the science" says. In any case, well, often an academic literature is just working from an oversimplified model, which may or may not be predictive, but which in any case is collapsing quite a bit of nebulosity.

## Epistemology

I'll be yet stronger here: I think if you're failing with respect to the standards of [epistemic rationality](#) I think you're almost certainly wrong. I'll discuss this more, but broadly speaking, if your beliefs make predictions that are generally or usually wrong, those beliefs are again [something like wrong](#). This can happen if you've made too strong of inferences on the basis of [weak evidence](#), or if you've just made stuff up.

And again, there have been debates about what counts as knowledge at least since ancient Greece, and while a lot of these arguments seem either confused or irrelevant, I'll claim that there are still some real questions here. A weaker point is that to some extent I think the debate about the many-worlds interpretation is really an epistemological question. My stronger claim is that I think we still don't really have epistemological theories that can account appropriately for [gnosis](#).

Briefly here: there are a number of different kinds of meditative experiences reported in different traditions, which are described variously as "waking up" (as though one's experience before had been a dream, or like sleep), or of awareness of an all-pervading consciousness, or also of knowledge of past lives, whether one's own or others. These experiences which are generally reported as revealing deep truths about reality or meaning, sometimes apparently straightforward and sometimes said to be ineffable.

To be clear, some of this I think is bs, and some are probably falsified in the first place. However, a lot of these I think are probably *legitimate*, but very confusing to grapple with or to know what to do with, either from afar, or personally (from my very limited experience).

## Axiology

This is where I'm most confused, and this is what most of this book is about. The [last chapter](#) of this book will discuss this all in detail, but briefly:

I don't believe that we have any *explicit* principles in common for understanding what goodness is, and I think there are real gulfs between the notions of value in use across the world.

Many cultures today think it's good and appropriate to beat children, to execute homosexuals, to practice self abnegation, to torture animals, to associate sexuality with taboo and shame, etc. etc. (Feel free to fill in whatever you would find horrifying, and conversely to fill in whatever *those* cultures would find horrifying about you.) There is a shallow modern liberal perspective which tries to ignore these sorts of differences in conceptions of value. I think this only works, insofar as it does, because immigrants mostly fairly quickly assimilate. I think most people don't quite understand cultural diversity and would often be horrified by lots of things that go on regularly in other cultures, just as they would likely be horrified at many things in your society.

A person, embedded in a given culture, might be bothered by specific practices, or might chafe at this or that hierarchy, but seldom do we have the time and space to ask "ok, but what's actually good? what's *really* good here, and what's actually just a mess?" I'll claim broadly that our current configuration has [very serious problems](#), and that it's very difficult to distinguish the *appearance* of cultural progress from status quo bias and cultural myopia.

But also, it seems like there's love and beauty and lots of both mundane and transcendent goodness.

I have a sense of what's probably the right direction here, but mostly a lot of uncertainty. Most of what I want to communicate in this book is a deep uncertainty about the nature of goodness, but without inclining towards nihilism. My thesis is basically, "there exists goodness, I feel very strongly about that, and still I'm quite uncertain what's really good."

## Communal metaphysical discernment

Of course, if one can't come to determination privately, how can we communally?

I actually have very little to say here. [I'll discuss more](#) about how I think ideas and values diffuse, and how they're negotiated culturally.

What I would add here is, I think that in practice most people are just kind of muddling through, and that in some ways this can allow for a kind of "communal discernment," as ideas diffuse through a culture, but that this is subject more to constraints about memetic *fecundity* (that is, reproductive fitness) than constraints about deep goodness or rightness.

## Footnotes

1. "resonance go brr," or something. [↩](#)
2. Thomas Aquinas's, in his [Argument from Degrees](#), argued that because beings can vary in perfection, there are degrees of perfection, and because (he sort of implicitly supposed) there must be a maximum, there must be a perfect being. IIRC Euclid's Elements contains a proof of there being no highest prime, it also seems trivial to just be like "dog but why does this set have an upper bound?" [↩](#)
3. This is usually glossed as "all mathematical objects exist, the physics and starting conditions of our universe being one of them." See the [Wikipedia page](#) [↩](#)
4. From *Through the Looking Glass* by Lewis Carroll:

[Humpty Dumpty] “And only one for birthday presents, you know. There’s glory for you!”

“I don’t know what you mean by ‘glory,’” Alice said.

Humpty Dumpty smiled contemptuously. “Of course you don’t—till I tell you. I meant ‘there’s a nice knock-down argument for you!’”

“But ‘glory’ doesn’t mean ‘a nice knock-down argument,’” Alice objected.

“When I use a word,” Humpty Dumpty said in rather a scornful tone, “it means just what I choose it to mean—neither more nor less.”

↩

5. FWIW I'm somewhat confused when the gender discourse came to be about non-binary people rather than binary trans people. It seems like the anti-trans side is upset about both but I feel like recent political actions (I suppose this dates when I'm writing this section) are superficially about non-binary people, and I'm not sure when or why that changed? ↩

# Rationality

---

This section explains a modern, basically orthodox model of rationality, and begins to introduce problems in its relationship to phenomenology, spirituality, and spiritual practice.

*In this section:*

First, a brief disclaimer on words, what they mean, what they are, and some unfortunate connotations for "rationality":

- [Words, and "Rationality"](#)

In modern theories of rationality, a distinction is made between two aspects: *epistemic* and *instrumental* rationality. There's a section here for each, and one inbetween on good epistemics:

- [Epistemic Rationality](#)
  - [Good Epistemics](#)
- [Instrumental Rationality](#)

I then discuss what I think rationality as a framework gets right, and the ways in which this perspective is circumscribed. I then claim there's a piece missing, and outline what the "third leg of the stool" looks like:

- [What Rationality Gets Right](#)
- [What Are the Limits of This Perspective?](#)
- [Spiritual Rationality and Rational Spirituality](#)



# Words, and "Rationality"

---

Some of this discussion is fraught, unfortunately, because of the bombast that the word "rationality" evokes. It's useful here to say some about what I think words are and aren't.

Mostly I would say, words are just that: despite the authoritativeness that the use of a word can have, words are "in fact" symbols which carry culturally constituted meaning, etc etc. This is to say, just because we have a word for something, or we're used to thinking about there being this or that category, that's in principle only necessarily a fact about the way humans see the world, and how that's reflected in our cultures and languages. Words can also refer to things which don't exist, or which group together many different things that actually have very little in common.

We can also define words narrowly in a particular context, and understand the word to refer to that definition in that context. Obviously, of course, there is still connotation, and a lot of abuse can be done by implicitly bringing in intuitions about some word or category, while [insisting that one is only using a very specific definition]. [\[\[link to old motte and bailey post\]\]](#)

With respect to "rationality" in particular, one would be reasonable to complain: "who are you to say that you own or can define truth, at everyone else's expense?" I think this is basically a good take, and insofar as I fall into speaking singularly authoritatively about "rationality," I'm mostly just being sloppy. *However*, I still think there's a very important collection of concepts which I need to make reference to... and it would only confuse matters to try to come up with a new name for them. Probably a productive angle from which to engage here would be to say, "ok, this is a fine enough concept, agnostic of the name," and notice what distinctions it makes salient.

Also unfortunately, a number of intellectual movements have used the word 'rational' or 'rationality.' I'm pretty thinly read here, and mostly pulling from Yudkowsky [\[link to sequences or R:a-z\]](#), but I believe the definitions below originally come from a literature on rationality in the 20th century, by way of decision theory, economics, and psychology, starting with Von Neumann and Morgenstern.

One last point: "rational" often evokes what Yudkowsky calls a "straw vulkan"[\[\[link to sequence post\]\]](#) —a caricatured person with no emotion, who only thinks in probabilities, weights, and causality. I think this image is less of a strawman than would be flattering, but I'll stress that rationality, as I approach it and as Yudkowsky presents it, is not fundamentally in contradiction to human richness and sensitivity. [\[\[this needs elaboration\]\]](#)

# Epistemic Rationality

---

*Epistemic*, from Greek *epistēmē*, "knowledge" <sup>1</sup>

Epistemic rationality: systematically improving the accuracy of your beliefs<sup>2</sup>

To begin with, I want to dispell some confused, popular concepts of "reality" and "truth," or at least contrast them with the notions that are used in the theory of epistemic rationality I'll be discussing.

In an ordinary sense, when we make concrete statements like "I'm sitting in a chair," there's a very real, if seldom salient, provisionality about perceptions and statements like these. Indeed, we experience and believe such things in dreams, and certainly people sometimes have stable coherent hallucinations that they later take back. Even if the truth or reality of perceptions and statements like these is seemingly just mundanely undeniable, they remain just shy of an absolute solidity.

That's of course the extreme case. The point here is that our epistemology needs to account for both fundamental human fallibility and fundamental uncertainty in any possible knowledge. (Perhaps, I should mention, this is the other side of such refrains as "physics is just a theory"—yes! Even in the orthodox theories of knowledge underpinning modern science, all human knowledge is uncertain, *however*, we should still have high confidence in this or that *theory*.)

The sense sense of "belief" that is used in the context of *Bayesian* theories of epistemology is very much like this. We gather observations, learn, and become more confident in this hypothesis over that one. **In Bayesian epistemology, there is no basis for making absolute claims about reality.** In this sense, there isn't an utter solidity even to claims like "I am sitting in a chair," rather these are regarded as beliefs with *extremely high credence*, or equivalently, beliefs which make predictions with *extreme probabilities*. This is the sense in which we can build back up to ordinary ways of speaking plainly about things like chairs and warm breezes and mountaintops, and *also* of claims about physics and evolutionary biology.

## **Bayesianism somewhat opinionated**

For what it's worth, my understanding is that Bayesianism is not accepted by all statisticians or probabilists, though I'm very poorly read here.

In practice I think this distinction is basically academic, I don't believe that other approaches actually generate falsifiable predictions in contradiction with Bayesianism.

In the jargon, one *updates one's priors* (essentially, learns and revises one's beliefs) on the basis of each piece of new information or evidence. "Updating" and "priors" may sound obtusely technical, but this is quite ordinary. Every moment we're learning, implicitly, quietly, just a bit more: a growing sense of what a friend is feeling, some particular detail about the life of a Japanese emperor, or just a bit better sense of the streets and shops in this part of town.

"Improving the accuracy of our beliefs" is then concerned with how to arrive at beliefs about the world which are *highly predictive*, as in, that they predict events with a high degree of accuracy and consistency.

This gets into what are known as *subjective probabilities*. I think most people find the concept of a probability basically intuitive when thinking about a literal dice roll, but when an event can only happen once, what can a probability even mean? In Bayesian probability theory, a probability is a measurement of the *confidence* in whether an event will occur. This works out the same for highly regular, but hard to predict objects like dice or jars of marbles, but also for whether there will be rain tomorrow. Suppose we assign a 50% probability to an event, eg. the outcome of an election. In this frame, that probability is interpreted to mean that, if we're calibrated, we expect for predictions with the same confidence to come true 50% of the time, on average.

Perhaps, imagine: your boss sends a message which is phrased in a way that could read as angry, or could just be hurried and poorly written. Whether your boss meant that text in a passive aggressive way and is in fact mad at you is not a *matter of chance*, but you can still be uncertain about it. If you assign 15% that he's actually mad, again this means that you expect to be right in 15% of situations where you feel similarly uncertain.

I want to stress here how the technical (and perhaps robotic-sounding) ideas of updates, priors, and predictions are actually mundane and natural. One makes updates whether one wants to or not, one can be more or less calibrated, and we live with the consequences of these every day, regardless of how we think about them. There are, however, domains which are comprehensible without any explicit theories of knowledge, *and* there are domains where one has to move carefully to come to correct beliefs, and that's what all of this is for.

Next we'll discuss what principles underly changing one's mind *well* and *poorly*, according to this frame.

## Footnotes

1. [OED](#) ↩

2. <https://www.lesswrong.com/posts/RcZCwxFiZzE6X7nsv/what-do-we-mean-by-rationality-1> ↩

# Good Epistemics

---

Traditional models of epistemic rationality end up constructing a *prescriptive* model of better or worse [updating](#).

These are placed to some extent within a specific [metaphysics](#), which I actually don't necessarily completely endorse. My take here is something like, "orthodox epistemic rationality is wrong so seldom that one should basically have these distinctions internalized, but nonetheless it's not flatly [comprehensive](#)." Maybe a more qualified claim would be "updates made this way lead to beliefs and models which are highly predictive, agnostic of one's goals or intentions."

It's also important to note that epistemic rationality is not a superstitious ritual purity. There are good reasons surgeons scrub in very specific ways, similarly why chip fab cleanrooms are maintained a specific way, and these don't (generally speaking) arise from superstition but from careful robust models, and thorough testing.

Note that while people *sometimes* do explicit probability calculations, in practice these are often used more like heuristics which guide ordinary intuition and reasoning, rather than constructing a tight algorithm.

The list below is hardly exhaustive but is both helpful in particular (and I'll be referring back to it), and also evocative of the broad flavor of what characterizes good epistemics, in this kind of model.

## Occam's razor

- what are our hyperpriors? what priors do we start with in general? why occam in general? solomonoff induction???
- I think part of this is like, I can get away with saying "look, this is a generative assumption, I'm not gonna justify it that strongly"

## Bayesian evidence and magnitudes of updates

- what is this actually evidence of, etc
- bayesian evidence is not legal evidence
- "strong evidence" does not mean dispositive, it means "inspires a large update" and weak evidence does not mean "a poor argument" or "probably false", it means "inspires a small update"
- base rates

## Parsimony of updates

## Privileging the hypothesis

## Beliefs pay rent

- this isn't the term that's used tho?

- correct magnitudes

# Instrumental Rationality

---

- there exist tradeoffs, there exist choices, there exists strategy: "instrumental rationality"
  - there's a big can of weeds here wrt game theory, optimization, etc; in this respect I'm also kind of shallowly read
  - symmetrically to what we discussed with epistemic rationality, again we can interpret this definition narrowly: the goal here is indeed to achieve our goals, to successfully get what we want (well, again, this gets hairy wrt spirituality, but we're not there yet)
    - sometimes you'll also hear "satisfaction of our values"
      - even avowed rationalists are like, yes! I have rich humane values! go cry and laugh and dance and do the worthwhile things!
  - naturally, there's going to be problems here. without even getting into awakening territory here, it's just, well, should you get what you want? isn't that harmful? what about colonialism, or rape, or who knows what?
    - part of the claim here, and this is a large part of the context of the understanding of values in a rationalist frame, is that "success" or something is not narrowly success in a competitive context, despite the language that's used
      - maybe we could define it, more richly, as "robust enactment of value"
        - yeah yeah but virtue ethics too, or something. obviously this is a frame
        - I think I'm sort of sympathetic to virtue ethics and I think I have a weaker commitment to instrumental rationality as a theory
        - I think, tho, that if you can't do these sorts of calculations you're going to just leave value on the floor, etc etc
- this definition is importantly agnostic wrt your values, this is either very clean, or presents problems if you're like "wait, but this is kind of degenerate and cold and wrong"--we'll touch on this next
- elaborate or refer back to "straw vulcan" here
- Unclear if I want to talk about newcomblike problems here. I can say something like "even in this domain we're having metaphysical debates"
- "effective" means "achieving the goals as specified"

# What Rationality Gets Right

---

- broadly just: if you're failing by these metrics you're definitely doing something wrong, but also this is not [comprehensive](#)
- what rationality get right: lots of stuff where I'm like, you have to be doing something metaphysically nuanced to say something that sounds like an empirical claim and is plainly false or not predictive. lots of people try to cheat here and I will continue to say that it's horseshit. so often they're doing a motte and bailey re like, "well, whatever, I don't care if this is empirically false"

# What Are the Limits of This Perspective?

---

- I think I want to express something like "these are great concepts, and they're also fraught! but, please make use of this concept!"
- something something this needs to be extended to be like, "I'm gonna build up to critiques of this more, but for now i'm just saying, yes this is a *purposefully* narrow perspective to make some technical distinctions. this is its correct use
- I'm not even mainly prioritizing what's wrong, I actually think this is great, please be familiar! and also, there's some problems
- Ordinarily, *most of the time*, we experience a separation between our beliefs about the world, and the world "*itself*."
  - this phrasing sucks but w/e, point is, I want to talk about like, ok there's hyperstition, etc etc, there's various contexts where this distinction isn't so flat
- what about fake it till you make it? what about, juicing yourself up? when is it generative to be *miscalibrated*?
- how do we update on gnosis? fucking who knows
- re limitations of rationality, we don't have really an account of hyperstition? Like, we do talk about this, so this is like "dark arts ", but we don't really have a way of dealing with this, we're just like dark arts, I don't know, seems questionable?
- Limitations of rationality: ppl do talk about moral uncertainty, and to some extent they talk about normative ethics but they don't really talk about transformative practice re all of this, and they don't talk about introspection re moral investigation

[ Some people end up reifying this sense of rationality, to end up saying that the only perspective or mental quality worth entertaining or embodying is one that results in accurate predictions. This is indeed fraught as a perspective and not what I'm saying. What we want to say here is that in regimes where the territory is completely or almost completely causally disconnected from one's perceptions and experience [ maybe just want to say, "when the map is quite separate from the territory" ], it's *very useful* to track models in terms of being highly predictive]

[ and indeed I don't want to use the word "true" or even the word "rational" in an abusive sense. "Reality" and "science" are so often used to browbeat people, as one way of constructing and wielding power. ] [ the use of science as a way of seeing that's exclusive of all others--this is not what I mean, and this is not the way that I make use of epistemic rationality as a concept.] [ link to ways of seeing post once I get there] [ but also, any set of goals is going to construct a way of seeing, you don't get a view-from-nowhere[prolly link that post?]]

There's a separate problem, and this is a large part of the relevance of spiritual practice: in human minds, and perhaps in principle, there is a feedback loop between priors, perception, [ and what updates one makes.] Existing priors [ determine how we categorize and update on incoming data anyway ] [ bounded rationality, etc, it's too expensive to keep recomputing your ontology, it's too expensive ] [you use the existing phenomenal construction to make contact with the territory, and you end up with, at best, a catalog, so to speak, of observations of the interface of the existing representations (priors) and the territory. a lot of what



spiritual practice is made to be about, is escaping feedback loops of specific perceptions, beliefs, affective structures, etc. ] [otuh ppl can overupdate in the direction of hyperstition, which is often harmful and to me pretty squicky. wrong practice, etc etc, the system wants to make good updates!]

- also here is like, what the heck is this reality thing? we don't actually directly encounter it, and what we think or perceive it to be is about what we already think and perceive it to be...

# Spiritual Rationality and Rational Spirituality

---

- 2025-02-28 almost certainly this stuff needs to be renamed

There are two very different concepts which I want to distinguish here clearly, hence they're on the same page: *spiritual rationality* and *rational spirituality*.

- I think I need better terms than "spiritual rationality" at least if I'm going to be distinguishing really 3 things: correct axiology, correct internal transformation (structure preserving transformations), and effective practice
- spiritual rationality: if you're engaging in a practice which will cause your values to change, or at least your understanding of your values, how do you do that right? What would it mean for it to be right?
- "the obnoxious move here would be to call this axiological rationality, but this doesn't really help us at all, nobody knows that word"
  - this also isn't just about axiological, so it's really wrt all of one's metaphysics
    - maybe the whole thing needs to be constructed differently? like "spiritual rationality" is confused?
- spiritual rationality is not just about correct transformation, but also just correct apprehension of goodness. Also I guess correct apprehension of "spiritual cosmology" (like if God actually exists etc, spirits etc, tho I'm mostly thumbs down on that stuff)
  - if I talk about "correct transformation" I'll have to link forward to some relevant section in practice, but basically here I can just be like "we don't want to throw out our values, so it matters that we don't self modify wrongly"
- somewhere I need to talk about metaphysical determination, maybe that goes here
- spiritual rationality is going to be like, ok, even if some major aspects of metaphysics are maybe impossible to come to determination about, there's still some where I'm like "no, that's wrong", which is what the rest of the book is going to be about
- look in "on attunement" from Carlsmith: "Give anti-realist rationality a goal, and it will roar into life. Ask it what goals to pursue, and it gets confused. 'Whatever goal would promote your goals to pursue?' No, no, that's not it at all." <https://joecarlsmith.com/2024/03/25/on-attunement>

- 
- 2024-12-19 sending messages in slack dm to myself about this voice notes:

[ so a big part of the relationship is like, spiritual practice is in many ways like, an unfolding of the nature of mind and experience, and indeed that should disclose things about the moral domain, like you do phenomenological investigation, and oh look, you find things about the mind! you find things about feeling and relationship and experience and indeed there's all these reports of big changes in outlook and

"worldview" and feeling and optimism and like concrete action and so on, downstream of doing spiritual practices but is that right? are we going crazy, is this stupid?]

---

## **Spiritual Rationality: the third leg of the stool**

in the first two legs it's taken as a given that one has values, and part of the claim is that in some sense epistemic and instrumental rationality as theories are agnostic wrt your values (this is a bit fraught, and both of them in fact make prescriptions that are contrary to a lot of people's intuitions in many cases)

this is maybe a fine stance if we assume that values are basically static; insofar as we're concerned with changing our *relationship* to phenomena, experience, etc etc then indeed we need a third leg here or the stool topples over! ["changes to the apprehension of one's values"]

## **Rational Spirituality**

so the problem here is not just, how do we evaluate changes to our values, changes to the apprehension of our values, we also care about doing it successfully!

# Culture, Trauma, and Memetics

---

*In this section:*

There's a wide range of phenomena that look and smell like "trauma," including a lot that are socially normative:

- [Trauma](#)
- [Trauma Smells](#)
- [Culture, Taboo, and Trauma](#)

Ideas and psychological patterns are the main transmission vectors of harm and disfunction in society:

- [Memetics](#)
- [Burdens of Meaning](#)
- [Discourse](#)

# Trauma

## ① Epistemic status: more like "my impression" rather than "established consensus"

This is also largely under-cited, and flat empirical claims are mixed in with some inference on my part.

The word 'trauma' is often contested politically, which I mostly want to sidestep. I'll explain first some context about the history of the word as I understand it, and then the cluster it's used to refer to in the psychiatry and therapy literatures. See also some context on [words and concepts](#) in general, and on [nebulosity](#) in particular.

There are some messy ontological and methodological problems in trying to analyze trauma. In contemporary rhetoric, 'trauma' often stands as a [disguised query](#) for a question like "who deserves privileges and accommodations because they have suffered." More deeply, whatever notion of trauma we use relies on some assumptions about what constitutes pathology and harm. Broadly speaking, 'trauma' refers to persistent *pathological* changes to psychology and behavior that arise as a consequence of adverse experiences, as well as the experiences themselves. That is, despite the language of "post-traumatic growth," positive changes that arise from adverse experiences are regarded as quite different from what we would normally be inclined to call trauma, and are treated as desirable, and even sought out deliberately.

## Defining trauma

'Trauma' in Greek literally means "wound," and is still used in that sense in physical medicine. [IIRC, Freud copied it from an earlier late 19th century psychiatrist, who I believe was following an earlier mid-19th century author, who wrote about theories of "hysterical trauma." The original theory was that a cluster of patterns of psychological disturbance called "hysteria", which were female coded at the time in Germany, were due to a literal injury of to the uterus (Greek *hysteria*, "womb").] [Again IIRC, the author who directly influenced Freud used the word in a more modern psychological sense.] [Freud's usage is relatively familiar to modern intuitions, though his actual theories are pretty foreign.] [I'm somewhat unclear what happened between Freud and the advent of the modern term PTSD.]

Because of the modern interface of medicine, law, and power, trauma is often understood narrowly in terms of psychiatric diagnostic criteria. There's some substantial confusion here, because the model which the DSM (the standard psychiatric diagnostic manual in the U.S.) uses is explicitly meant to track clusters of symptoms that will likely be tractable to similar interventions, but not necessarily to track natural kinds, let alone sharply defined causal mechanisms.

The DSM and the ICD (the International Classification of Diseases) use very similar definitions for PTSD, and have a very similar list of adjacent stress or trauma-related disorders. While I don't take these definitions as singularly authoritative, I think they're informative as models developed by people earnestly trying to describe and work with this domain, as well as carrying quite a lot of authority, for those who put stock in that.

The criteria for PTSD approximately match the stereotyped image of victims of war, genocide, rape, etc. These involve, briefly: a. flashbacks, intrusive memories, or reactions to associated triggers, b. avoidance of stimuli associated with the trauma, c. persistent emotional distress, alienation, or anhedonia, and d. persistent emotional dysregulation, stress, and fear.<sup>1</sup> These may be attributed to acute intense disturbing experiences, or more chronic and sometimes less intense.

The ICD (but not the DSM) also has an adjacent category known as *Complex Post-Traumatic Stress Disorder*, which involves the same criteria as PTSD, plus distinctly strong feelings of shame and worthlessness, and severe difficulty sustaining relationships.<sup>2</sup>

(Note also that therapy organizations which have been advocating for the diagnostic status of CPTSD believe that it's something like "responsible for" a large swath of very specific diagnostic categories in the DSM. IIRC the definition they use is different and less restrictive than the one in the ICD.)

Even in the ontologies of the DSM/ICD, there are several other categories of trauma-induced disorders with different symptoms, which are somewhat weirder and less stereotyped, eg. Reactive Attachment Disorder and Disinhibited Social Engagement Disorder. Note again that the ontology of the DSM is constructed in order to identify clusters that will be amenable to similar sorts of treatments, often just on the basis of their overt symptoms, rather than necessarily in terms of their underlying causal structures.<sup>3</sup> The resolution at which psychiatric treatment is applied broadly speaking can't track the underlying causal structures, which might be unknown, highly multi-factored, or highly varied. The techniques of psychiatry are often very "try stuff and see what sticks."

"Psychological" as opposed to "psychiatric" models often revolve around detailed inside-view theories of the causal factors responsible for dysfunction, which are empirically much weaker, and which psychiatry usually avoids. (This is sort of a simplification, my understanding is that psychiatric theory does include these kinds of sketchier inside-view models also, even if diagnostic manuals avoid them.) Note also that "psychological" models are willing to use far more flexible admission criteria, rather than psychiatric models, whose diagnostic criteria are often defined in order to have thresholds at large magnitudes of impairment. This is to say that diagnostic criteria often have more to do with contingencies of the modern medical-legal-insurance context, rather than reflecting some sharp boundary in the domain.

In these slightly-less-orthodox models of psychopathology, trauma is regarded as an important causal factor in (and pretty straightforwardly correlated with) a variety of different disorders, eg. Anxiety and OCD.

This is all a very long-winded way of saying, "trauma is not restricted to just whatever the DSM says."

There's obviously then, a big question: within the broader category of "psychological changes that come as a result of adverse experiences," what intuitions can we appropriately apply from the more clear cut "trauma" of the diagnosable-PTSD sort, vs. milder and subtler kinds of changes, in particular re whether they are desirable, and whether they should be regarded as a kind of harm. I'll refer to this question as "what should count as pathological?"

For now I'm going to leave this question up to the judgement of the reader. I'll try to just describe what's actually in the various reports, anecdotal, and actual data, and I'll discuss next [a particular model](#), and then [trauma smells](#). I'll discuss this question again partially vis a vis [introspection and transformative practice](#) and in detail [near the end of the book](#).

## Models, reports, and anecdota

All of the components of the diagnostic criteria for PTSD (and I left out some detail) are reported in various combinations and severities due to adverse experiences. There's also a variety, both in the normative psychiatric models/ontologies and more "extensive" psychological models, of other kinds of "trauma-induced" pathologies that don't just look like PTSD. Some broad clusters here include exaggerated (to extremes) neuroticism, self-centeredness, neediness, etc.—basically ordinary, if undesirable, tendencies in human psychology in general. Again, there are milder forms reported that seem to be caused by traumatic experiences, which don't necessarily qualify by official criteria.

There are three important (non-exhaustive) clusters of traumatic symptoms reported in the therapy literature (note these are not quite standard terms):

- **Distortion:** patterns of variously intense kinds of psychological "lopsidedness," dysfunction, or "distortion" which are claimed to result from abuse or neglect.
- **Rigidity:** an intense need to maintain one's behavior, appearance, or sustain positive regard from others in very specific ways, for fear of rejection or some projected memory of violence
- **Dissociation:** numbness, dissociation, and deficits in emotional development

Therapy manuals are rife with reports of these sorts of cases, of course aligned very nicely with the views of the given therapy modality or school. The stereotype of at least a modern genre of these stories usually goes something like, "I realized that my exaggerated, dysfunctional, neurotic behavior was the result of unresolved trauma. As a child [*and the stories are almost invariably about childhood trauma*] I learned those emotional and psychological patterns as a defense mechanism, but now I am able to heal that trauma."

I'm sort of uncertain quite how to take this evidence. It seems to me that schools of therapy often have dodgy epistemic standards, and my impression is that the best "evidence based" therapy actually doesn't work all that well, in the sense of producing robust recovery, etc.<sup>4</sup> One plausible model for what's going on here, which I like, is that the effectiveness of therapy is contingent on extreme levels of skill and intuition on the part of the therapist. This level of skill would be difficult to transmit, measure, and document carefully, leading to a pretty severe regression to the mean in each generation of some school of therapy. See [more on this subject](#). Nonetheless this is decent evidence against the accuracy of these models.

Still, my sense is that this broad shape is probably substantially right. Some evidence that contributes to my sense here:

- IIRC the correlation between eg. ACE (Adverse Childhood Experience) scores and dysfunctional adult outcomes is pretty strong. Similarly, again IIRC, for other quantitative measures of the correlation between trauma and dysfunctional behavior.
- We have lots of quite tidy evidence about the structure of trauma responses in the extreme cases (actual PTSD), that traumatic psychological adaptations are something like "specific" and "structured," (as in, specific to the details of the trauma, and shaped in a way that reflects it) rather than resulting in someone being only diffusely disturbed.
  - It seems basically plausible that we shouldn't expect this to be specific to extreme cases, and while the evidence from therapists is sketchier and less robust, it still seems pretty substantial to me.
- This broad claim, that adaptations to traumas are structured with respect to the shape of the trauma, aligns with my own introspection and with introspective reports I take seriously.

- Note that I think most introspection is mostly nonsense, see [more about introspection and confabulation](#).

In the broader definitions used in the therapy literature, trauma is not just restricted to experiences of intense fear or horror, but also often of experiences of grief and mundane tragedy. Commonly reported as well are chronic, low to medium-grade experiences of alienation, rejection, humiliation, etc.

Reports of the dissociation cluster often come with healing stories, to stereotype, of discovering some great pit of fear, anger, sadness, etc., which the person reports they were "protected from" or which was "held at bay" by the previous dissociation. They then report "processing" those emotions, and recovering the ability to experience some greater breadth of emotions, body sensations otherwise, or of newfound ease and comfort with social interactions.

These reports also usually include variations on the some themes:

- Recollection or "discovery" of like, something in the genre of the breadth/depth of grief from some incident or chronic adverse experience, and of contortion of their psychology wrt this
  - Commonly, recollection of child abuse, long suppressed
- Recollection/discovery of unusual or proscribed proclivities, stereotypical stories here are from trans people, queer people more broadly, from the BDSM community, etc.
- Less commonly, discovery of "held" grief/fear/etc. around the location of a previous physical injury, chronic illness, or body part associated with past embarrassment or shame

The broad thrust of these reports, both in the very modern modalities but also going back at least to Freud, is of discovering a lot of structured, highly "semantic" psychological content associated with gross psychological dysfunction or "offness", exaggeration, distortion, etc.

A similar cluster that's often reported includes patterns like tendencies towards hyperfocus or obsession, the apparent absence of normal desires or appreciations (this overlaps with dissociation), compulsive nonconformism, unusual superstitions, etc.

Again, this entire category of evidence is fraught. Many of these reports observe patterns similar to [culture-bound disorders](#) where people have experiences of neurosis, grief, etc. in line with very *specific* patterns in their culture, and often in accord with some changing zeitgeist of eg. western therapy culture. Therapy reports are also problematic because they're heavily primed, even where a therapist is skilled and not setting a narrow frame, clients often come in with a particular image of what healing "should" look like, and will enact that. [I'll discuss later](#) my sense of what constitutes *real* or *trustworthy* introspection, though broadly this is pretty hard to justify in an "objective" way. Still, I think as a broad shape these are very real phenomena.

Introspective reports from solo-practitioners, meditators, etc., including anecdotes I've collected, overlap substantially with those from the therapy literature, but more varied and probably more dubious. Again I think these are *often* not trustworthy, but are an important source of evidence as well. Some of the weirder stuff here is also actually attested in the therapy literature, eg. encounters with demonic entities which entered one's psyche during a traumatic event, so-called "genital armoring," or experiences of accessing ancestral or past-life traumatic memories.

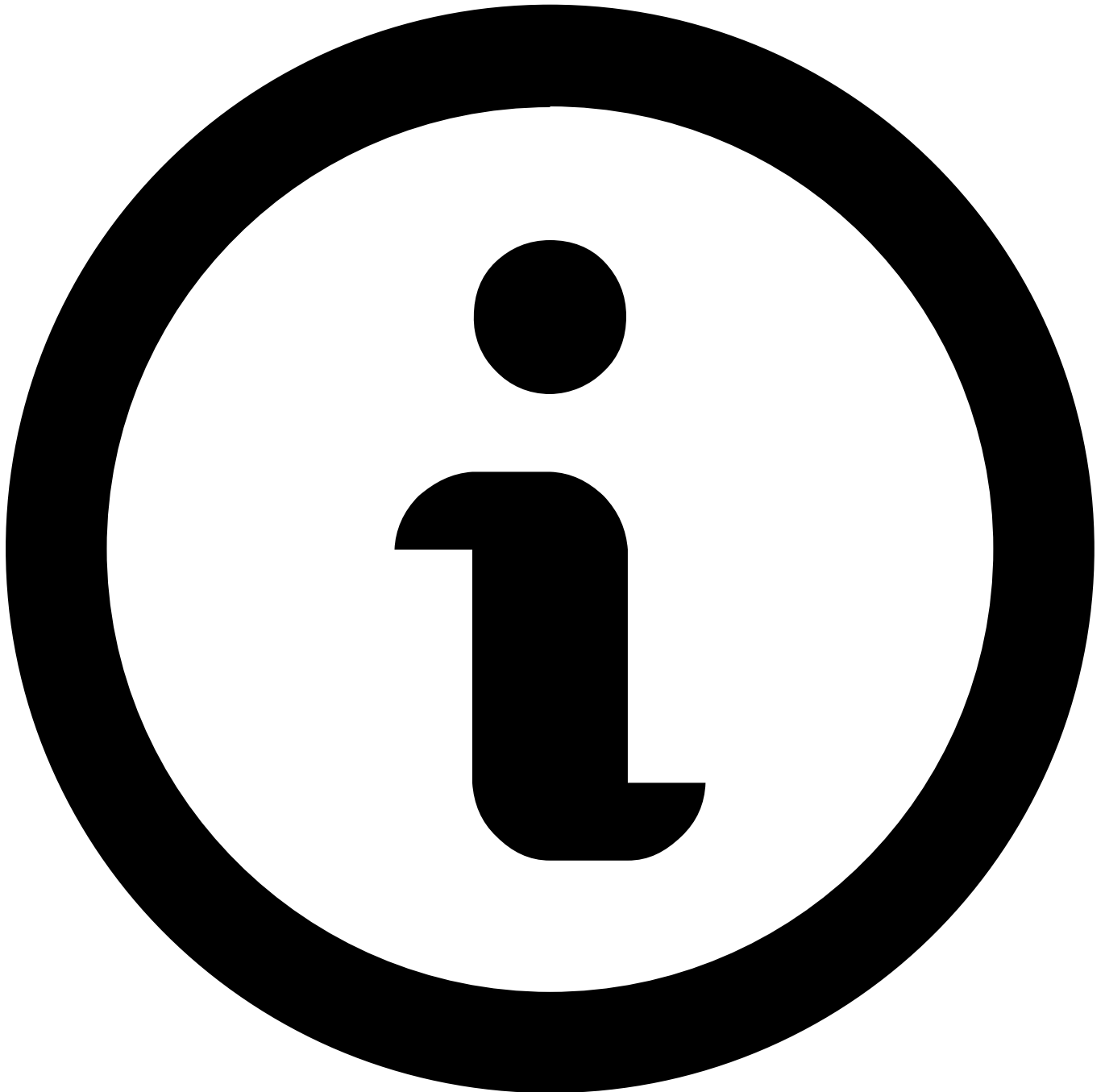
## Footnotes



1. American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*, pp. 271-280 ↩
2. [[cite icd <https://icd.who.int/browse/2025-01/mms/en#585833559>]] ↩
3. [[prolly cite scott here talking about the dsm]] ↩
4. ↩

# The Structure of Trauma

---



*Epistemic status: speculative, with some meaningful evidence*

The image that one might get about trauma based only on the criteria for PTSD is that it's mainly about symptoms of emotional *dysregulation*. I think this is too narrow, especially for trauma smells. One view that I

like says that trauma is something like an *exaggerated* and overly restrictive response to real fear, horror, and grief.

It seems like in general, mammals, at least, learn on a kind of "emotional" basis, as in, following gradients of valence and emotion. This often follows some thread of curiosity, satisfaction, or pleasure, or their opposites. This can be gentle and slight, or can also take the shape of very large [updates](#) in response to acute, intense valence, whether positive or negative. IIRC both of these are observed pretty straightforwardly in primates, dogs, rats, etc. Perhaps, some people are capable of learning in a "neutral," "technical," or "clinical" way from experiences that for others are strongly valenced, but descriptively this kind of emotional imprinting is how things usually go for most people, for experiences of extreme danger, joy, etc.

Organisms in general also need to be highly sample-efficient when it comes to danger: you don't get many chances to almost die, or almost lose your leg, and you want to *really* learn your lesson from near misses. Humans can also transmit various kinds of dangers culturally, besides the much slower and coarser transmission of genetically encoded instincts that lots of species have. Still, cultural transmission is generally pretty low-resolution compared to the full reality of some threat, and even if people have internalized that something is dangerous, when they actually suffer some harm they likely won't be "prepared" for it emotionally.

So then, one updates hard, and often one both [over-updates](#), and updates in a way that becomes very difficult to roll back.

Besides this, there is also so-called "negative trauma," basically trauma that occurs due to the *absence* of some experience. The prototypical case here would be childhood neglect. Now, "neglect" usually correlates with, and often has a connotation of, other kinds of adverse experience, but there seems to be a real effect where a child's environment is not overtly abusive, but there's an absence of parental warmth and care which leads to both "psychodevelopmental atrophy" as well as patterns of negative emotion etc. more stereotypical of trauma.

## Blending with the view

Besides just for sample efficiency, updates about negatively valenced outcomes probably "ought" to influence our feelings and [ways of seeing](#). While spirituality has some bearing on this, under normal circumstances it's going to be something like "appropriate" to temper one's sense of rightness, comfort, and emotional safety with respect to the real circumstances in the world. Barring the far reaches of [spiritual attainment](#), if you know that your brother is suffering, or that you might lose your job, this probably should be integrated into your affective sense of the world. Again, descriptively, this is usually how things go.

There's a tendency for these emotional updates in response to traumatic experiences to get deeply ingrained in the default ways of seeing, often in ways that become all-but-invisible. Reports from meditators, who often have richer introspective access than most, often go something like, "yeah it just *felt* like the world is fundamentally horrible and unsafe, and that colored everything I saw and interacted with"—and many such variations. Again, these reports can come from survivors of genocide, but also (and sometimes with similar language) from those who've experienced totally mundane personal tragedy, familial estrangement, etc. A very stereotyped example here comes from victims of bullying over many years, who report that their experience of being low-status, shameful, etc. seemed to be simply "part of the world," or so much a part of the world that it was invisible until the burden sloughd off.

I'll refer this kind of phenomenon as cases where trauma is highly "*blended with the view*."

PTSD proper is partly also about when a person sort of keeps "triggering themselves," but that's less what I'm talking about here. There also seems to be something especially pernicious about negative valence, where it's self-sustaining in a way that positive valence generally isn't. Maybe, we can view [awakening](#) as an experience that's *so positive* that one never needs to feel bad again, but this at least isn't what happens to most people when they fall in love, or win the lottery, or whatever.

## Emotional/Psychological Metabolism

Related to all of this is the concept of *metabolism*. For most kinds of negative experiences, and even sometimes for very intense ones, there's a variety of ways that the experience/emotions/"content" etc. can be [integrated](#). Two tentative glosses on "integration" here might be: a. robustly and maturely understood, and weighed appropriately with respect to the rest of one's values, and b. no longer acting out a traumatized, injured response, having a mature relationship to the reality of some experience.

So, let's say, a child experiences a let-down and is utterly distraught about this, but with time comes to a more mature relationship with success and failure, at least in degrees. Not to make a blithe comparison, but probably the same is possible with respect to loss, safety, humiliation, etc. There are familiar stereotyped reports about "learning from" the loss of a loved one, of a kind of maturity and richness and flowering that comes from it, or through it, or whatever.

Part of the experience of metabolism, in the caricature both of childhood let-downs and of gut-wrenching grief, involves some intense emotionality, tears, "catharsis," etc. This accords with the reports I generally buy, whether anecdotes from joe-off-the-street, meditators, and from therapy manuals. It's salient here to note that adults, and stereotypically, especially men, often have lost this kind of catharsis as a regular feature of their lives. Perhaps, sometimes this can be because they're already "highly integrated" in this domain, but from my perspective this is an indicative trauma smell.

To be clear, a lot of what looks to me like "correct" or "robust" metabolism happens on its own, and is not necessarily loudly emotive at all. There's some cluster of evidence that trauma, and maybe especially "small-t" trauma has more to do with not feeling safe socially in a negative experience. I'll [discuss more](#) about subtler and more unusual kinds of metabolism.

There are very common reports of low-grade chronic anxiety, malaise, ennui, and even gross psycho-social dysfunction "unraveling" immediately causally due to intense catharsis. Just the same, reports of healing experiences from both "classical" and "small-t" trauma do often actually look pretty much like stereotypical cathartic metabolism as well. There are however reports of kinds of very advanced metabolism that are "rarified," and might be silent or invisible from the outside.

## "Correct" metabolism

See also some discussion of [correctness, in general](#).

This all leaves out what the "goal" or "resolution" or "end state" of metabolism ought to be. This turns out to be a larger question, which I discuss as it relates to [psychology](#) broadly, [spiritual practice](#) in particular, and [meaning/value](#) in general later. Still, I'll discuss what some of the considerations are, as it informs what I consider to be trauma smells.

First, there exists real suffering, violence, and injustice, besides which at least mortality seems to be basically inviolable. Most people are by default only sort of *vaguely* aware of the reality of suffering that they haven't experienced, and it seems to be difficult not to come away disturbed from facing horrors up close. Even very "personal" trauma of humiliation and abuse might be less profoundly disturbing just to witness but is still quite [distorting] to actually receive.

It's obviously not sufficient to say "well, those who haven't experienced such suffering represent some kind of *truer* untraumatized state"—there's real suffering which such a person hasn't really updated on. It seems like we want some notion of "maturity" or "integration," something to the effect of "wisdom despite, or through, suffering." So the question is sort of, what does maturity look like, and what is "false" maturity? (Or perhaps, trauma is just a ratchet, and you can only be dirtied or injured by it.)

A traumatized person has in some sense a thesis about what is safe to do, think, or feel. Agnostic of whether this is "correct" or "appropriate," broadly speaking it's in any case an adaptation to some adverse experience. An integrated perspective has to meaningfully take into account the full reality of the adverse experience in the first place, and not just by "faking it" and trying to [say magic words](#) which are not *existentially real* for you, nor by [dissociating](#), or by just ignoring the subject.

So again, mortality. As far as I can tell, you and I are both going to die. If you learn your death is imminent, just learning that will probably be quite intense and painful, but you ostensibly already know that's coming. Have you faced that? What would it mean to have faced it? There are lots of spiritual/existential perspectives about death which largely look [wrong](#) to me. Again, I'll [discuss later](#) what a "correct" perspective would have to include here.

# Trauma Smells

---

I'll use the word "trauma smell" to refer to a big cluster of things which, well, *kinda smell like trauma*. The term is a (shallow) riff on a concept from software engineering. [This term is useful to refer to a lot of stuff that looks like trauma, but which is both uncertain in particular and also [nebulous](#).]

I argue for a few claims here, some basically orthodox/normative and some disputed but fairly common in the circles I run in:

- Trauma is the result of unmetabolized adverse experiences (though not all such result in trauma).
- There are many milder forms of trauma, and there's a broad swath of "trauma smells" that look like trauma.
- There's a smooth gradient between both the causal structures and symptoms of PTSD proper and a lot of things which constitute normative socialization in many cultures. (See mostly the following page on [cultural trauma](#).)

I've [already said](#) that trauma is [nebulous](#), but still I'll talk about some distinctions here.

At minimum, I probably want to reserve the word 'trauma' at least for phenomena that are caused by some experience, rather than due to something like underlying genetic variation. It's then often unclear what we're actually looking at. There's quite a lot of variation which we still probably want to call "pathological," both in medicine outside of psychiatry, and in psychology/psychiatry. "Genetic" variation and adverse experience also interact, of course.

More broadly for pathology, again nebulosity here. So, is aging pathological? Again, see [disguised queries](#)—to a large extent this question is either unhelpful or irrelevant, but it's fine to say "well, this seems bad, and it has some causes, and it's preferable not to have it," and leave aside some mostly meaningless metaphysical question. In any case, re trauma smells, there's quite a lot of stuff that sure looks pathological, and which looks like it has its cause in adverse experience. Partly also, "trauma smells" doesn't quite rely on assumptions about the etiology, mostly I'm saying "here's a big cluster of things which smell bad, and which seem to be pretty common."

[As I said](#), the word trauma often stands as a disguised query for "what deserves special privileges," but I'm not trying to make some rhetorical move by using the word here. That's part of why I'm using this silly word "trauma smells," and that's also why a lot of my friends often literally use the word 'stuff' rather than try to litigate this or that as trauma.

## Smells in particular

First, two methodological points here. One, these are in general something like "heuristically informative," or maybe just only weak to moderate evidence. Partly they're also a useful cluster to be able to talk about, regardless of how strong of evidence they are for "trauma proper," or whatever. Secondly, these mostly come

from vibes-level, "energetic" intuitions on my part, though not just mine, and I expect my meditator friends are mostly nodding along at this section.

There are [three main clusters](#) of traumatic symptoms and trauma smells, namely, distortion, rigidity, and dissociation. I'll weave in and out of talking about each of them.

One grouping of trauma smells, which cuts across the three clusters: how readily is a breadth of [ways of seeing](#) and of ways of being available to them? What kinds of psychological factors and states do they have access to? What seems to be forbidden to them?

- Do they have access to a breadth of both positive and negative affect?
  - to ease and relaxation?
  - to dreams and hopes?
  - to excitement and exuberance?
  - to devotion and sacredness, agnostic of the source?
  - to love and intimacy? To sexual arousal?
  - to feel fear?
  - to feel anger?
- Do they find it okay/safe to behave in a full range?
  - to sing and dance?
  - to dream and to hope?
  - to *be* angry?
  - to be loud?
  - to be harsh and unkind?
  - to *show* fear?
  - to feel or show weakness?
  - to feel or show strength?

These are all very common and fairly indicative trauma smells.

Often when something is "forbidden" to a person, rather than them explicitly saying "oh I can't do that," "I mustn't do that," etc., they will encounter internal resistance when they incline towards some way of being, or they will find that domain of meaning blank, empty, irrelevant, or they'll say that they "never got it," "never experience that," etc.

Now, there does seem to be some [natural variation, but etc etc I'm usually more like "nah dude that's a trauma smell"]

[And, like, broadly speaking, if someone says I do not experience this, like, yeah, there's some evidence that this is, like, you know, there's the, like, FAAH out, blah, blah, blah, and so it's, you know, I think that there's, like, some evidence for something like, yeah, there are some cases that actually it's just, it just doesn't happen.] [And then there are reported cases from, like, mystical traditions, wisdom traditions, et cetera, that are like, oh, yeah, indeed, you can purify it.] [But definitely among normal people, if you find, oh, I never experienced that, that looks like a trauma smell to me. That's a very indicative trauma smell.] [that's one cluster.]

[Oh, also, like, like, noting when someone is sort of, like, seemingly, like, like, you know, like, I have a sort of intuition that it's, like, there's a sort of, like, breadth of human meaning that, like, broadly speaking, should

work for most people.] [And then when you notice, like, okay, well, this person, like, seems to, like, be, like, restricted in the breadth of meaning that they're sort of, like, able or willing to access, that's a trauma smell.]

[some things that look like, you know, uh, this, this sort of domain of meaning being unavailable is more, like, it's sort of, like, foreclosed that, that meaning will only, you know, like, give rise to negative valence or will only give rise to negative valence in the, in the long term.] [And so, like, that might appear as it just, like, there might, like, grossly appear as it being unavailable or them, like, not partaking in that domain of meaning, a person not partaking in that domain of meaning, but then you actually, like, if you sort of, like, really poke at it, like, or they, like, disrespect or whatever, it's, like, oh, yeah, this is horrible.] [This is evil. Uh, you know, this can only lead to bad outcomes, things like that. ] [Very, very common cluster there.]

[it's not just negative, it's also positive.] [it's not just, I must not feel this way, but also I must, I must feel this way.]

- So I have to have positive affect.
- intense need for things in ways that seem out of whack. (distortion)
  - intense need for approval
  - intense need for reassurance
  - very intense, like, sexual something-something. (Now, obviously, like, a baseline here is, like, to be clear, like, young horny people are quite horny, this is fine, but also hypersexuality is a very strong trauma smell)
- effort or energy needed to sustain some way of seeing, affect, relationship, etc
  - "it requires effort to achieve this at all." [Like, I have to, like, consciously, purposefully, blah, blah, blah in order to have access to it.] [Now, it's more complicated. I think there's more nuance there.]
- demands for things about gender, very common.
  - this can be variously extreme. maybe this is a widespread cultural trauma, but also to some extent, shy of a maximalist sense of trauma, it's like, well humans seem to want to be gendered-conformist by default etc.
- Demands for things about class, very common.
- somewhere here I'm going to talk about energy signatures

[unclear where this fits in but important: ] [So, oh, yeah, so, like, some of the structure of certain kinds of obsessions are, like, one kind of trauma smell is, like, a certain kind of, like, relationship to some obsession that looks like there's a kind of, like, very tight knot of contraction with respect to some unresolved grief that has, like, you know, contracted itself with respect to a particular, sort of, like, object level goal or object level fantasy or sort of, like, you know, affective relational fantasy, and that, like, it's sort of, like, unable to update.] [And, like, inability to update, that's an important kind of trauma smell.] [That's, like, that's, like, that's, like, that's actually, like, weak evidence, but that's definitely, like, that's a kind of trauma smell.]

[So must doesn't necessarily mean that people will insist I must.] [You do definitely find that, and that's, like, a very strong trauma smell.] [But also, like, something like unavailable.] [Unavailable, impossible, inconceivable, you know, opaque.] [Of all of these are, like, you know, indicative trauma smells.] [or, some



affect or way of seeing is always online despite what sort of naively "should" be present, other feelings that they really must be feeling, or ought to feel, or whatever]

[You know, along here also is, like, reported in, like, you know, from various kinds of meditators, practitioners otherwise is, like, oh, I did some meditation, and then, you know, I did all this meditation for a while, and then a thing which was not available for me, which did not exist for me, and I thought was very neutral, I actually discovered was there was some, like, structure, there was some layering.] [And then now I have access to it.]

- prolly have to link forward about layering

## **energy signatures/fields, vibe reads, etc**

[so those are kind of, like, gross properties that you can notice, like, pretty abstractly as, like, interesting things.] [Now, what's then more interesting is that if, like, go and you look at, it's also, like, energetic, sort of, like, energy signatures, right? So, like, you look at people and you notice when they have those properties, like, what is the quality of their aura? ] [what energy signatures that are correlated with the, like, gross properties that I described here?]

- obviously have to introduce this in an anodyne way
- will obv have to link forward about energy in more detail in the practice chapter
  - briefly, like, okay, stuff about energy, stuff about blockages. [And then I'll just say like, here's some things that this often, this can look like and there's some things that the internal phenomenology can be like.] [And so that can look like blankness, can look like contraction.]

[And, yeah, there's, like, a bunch that you, that you notice and I, there, this is, like, harder to characterize and, like, mostly not possible to justify.] [Like, maybe I can just say here, like, yeah, this, like, accords with my perception, but, like, my perception aligns with this.] [I probably am getting false positives., I'm probably also getting false negatives.] Um, probably when I talk about energy signatures, I just want to be, like, yeah, it's, like, you know, like, there's a kind of, like, affective blah, blah, blah, that, like, maybe there is, like, some abstraction here that's possible (meaning like, I'm still basically a materialist here, and maybe you can refer to a bunch of gross properties that do just show up on a camera. tho also they could be very fine, very subtle, and you're just picking up on a lot of stuff), but, like, you're going to experience it in a sort of, like, simple affective whatever way that is going to, like, phenomenologically occur sort of, like, according to, like, energetic whatever.

- fragility.
- Kind of, like, energetic thinness.
- how much fear does there seem to be?
  - Fear and activation.
- exaggerated played up charisma and like, warmth or humor or something.
  - [sometimes that's conscious, and sometimes people just simply are trying to manipulate you.]
  - more interesting, and common: does there seem to be ease in it?
    - Does there seem to be activation in that quality?
- Or is there a kind of resistance in it?
  - Is the system kind of, like, fighting itself?
- ease that seems in contradiction with other things.
  - okayness that doesn't seem to be okay.
  - Relaxation that seems to be tight.

- kink in terms of distortion.
  - So, obsession with a particular point, with some particular input, um,
    - things that probably look like enacting a trauma.
  - [it's a little unclear because it's, like, there seems, and I'm going to talk about this way later, but, like, there definitely seems to be evidence for, uh, something like some kind of, like, underlying, um, sexual, like, like a very, very gentle DS.]
    - [Like, it seems like DS is basic, but it's, like, hard to, it's hard to, it's hard to quite tell, uh, because, like, um, you know, like, uh, something, something patriarchy, something, something, like, you might imagine that the sort of, like, expectation of DS in a, in a relationship is in some way sort of proportional to the kind of, like, strength of patriarchy in that culture.] [I don't know if that's true, but that's, like, that's, like, a story that we might tell. Um, and it's a story that certain kinds of feminists, I think, have told.]
- exaggerated social submission, exaggerated dominance and exaggerated submission.
  - [there is, like, local deference, there is local power. sometimes this smells off, but I don't rly want to call it a trauma smell] [there's probably something there that's appropriate.] [at some level sort of, like, psychologically appropriate of, like, both how you sort of, like, like, like, probably, like, how you show up energetically in terms of dominance and, just, like, ordinary mundane, like, dominant, you know, sort of, like, yeah, dominance and submission.] [But also, like, how you show up, um, um, uh, like, what, what claims you make, how you sort of, like, particularly participate.] [So it's energetic and it's sort of, like, grossly behavioral.]

[this list isn't comprehensive, that's fine etc]

- maybe also like, what does the opposite of this look like? groundedness, richness and sensitivity in positive valence (not just exuberance), richness in negative valence.

[And there's, like, extremely fine grades here, which I don't, which I really don't want to characterize as trauma smells.] ([the point here being like, well, they might just be scared at the moment. none of this is to pathologize fear, activation, anxiety, smallness, etc etc per se] [but we're trying to track their baseline])

## what informs categorization as a trauma smell?

[again, on vibes, as I said above]

[it's not just to say, like, well, untraumatized people sort of, like, totally have it right, but if you look at untraumatized people and you sort of, like, notice what sort of properties they have, and also if you look at, like, seemingly less traumatized cultures, and you go, like, okay, well, what things do they have there?] [I think it's reasonable to say, well, in some sense, that's a kind of, like, you know, there's a kind of, like, ignorance is bliss, right?] [part of the reason why they're sort of, like, energetically cleaner is, like, kind of ignorance in it, but, like, that doesn't necessarily defy my notion of, like, okay, again, this is a kind of trauma smell. ]

[Now, the claim here is not, like, okay, you need to, like, look at the median person, and the median person is untraumatized. But you look at the median person, and it's informative. And then, but also, it's sort of, like, you look at what clusters exist that are correlated with the, like, sort of traumatic load in a culture. ] [And you go, like, okay, if there's, like, a context where there's less traumatic load, like, that's informative about, like, what, like, a relatively untraumatized psychology looks like.]

- prolly link forward to traumatic load
- something like "each of these is heuristically informative, most of them aren't definitive"

[well, maybe they're just neurodivergent?] [ blah blah natural variation] [obviously a lot of my claim here that this is, like, messy and a little hard to tell is, like, and, like, but I'm just saying, like, this is what seems to accord with, this is what a lot of my perception is, is, like, there's a bunch of stuff that's, like, now claimed to be neurodivergent.]

[And, like, the claim of neurodivergence, of course, is, like, oh, this is just sort of natural variation. This should be respected in the same way that, like, left-handedness is sort of neutral, is, like, basically, like, completely neutral.] [the thing they're trying to sort of argue for is, like, oh, this shouldn't be regarded as pathological.] [This is just natural variation.]

[And I'm, like, well, a lot of this actually looks like trauma smells, to my perception.] [partly bc of introspective evidence as it gets healed etc. ]

[also something about the memetics of pathology. like, you end up making the thing bad and shameful, etc.] [and a bunch of contemporary movements are about trying to remove that stigma, which, maybe is fine, but they're like doing so by just determining by fiat that there's no pathology there]

- probably link forward like, "trauma smells relates to tech debt"

# Memetics

---

A **meme**<sup>1</sup> is a piece of cultural information, and memetics is their study. Memes include everything from written language, to manufacturing techniques, individual songs, and religious rituals.

Mememes have properties that make them similar to organisms, in particular to a variety of species that require hosts. These include parasites, bacteria, viruses<sup>2</sup> and even technically mitochondria. Host-dependent organisms may be helpful, harmful, or neutral to their hosts.

Mememes reproduce, mutate, and are selected like ordinary biological organisms, though through different mechanisms. Mememes are also subject to different selective pressures from biological organisms. Mememes, of course, will die out if they don't reproduce. A meme's fitness is determined, among other things, by its transmissibility, persistence, the effects that it has on its host, and by how it competes with other mememes in its niche.

## Memetic survival and reproduction strategies

In particular, there are a number of strategies available to mememes, like there are to biological host-dependent organisms:

- Gut microbiota are ~impossible to get rid of without modern antibiotics, and are *helpful* to the fitness of their hosts. They are usually transmitted by mothers to their children, and are about as resilient as their host populations are.
  - Mememes like "shoes" or "agriculture" or "money" probably belong here.
- Some kinds of infectious agents are *endemic* within a host species. This can either mean that most members of the species are infected for most of their lives (like HSV-1, which >50% of adults have) or that the agent moves around a larger population, being eliminated by individual members but often returning, like the common cold.
  - The former is maybe comparable to superstitions, and the latter maybe some kinds of harmful dieting fads.
- Some kinds of infectious diseases often kill their hosts, but are highly transmissible, eg. bubonic plague, maybe less so COVID.
  - Probably suicidal terrorism isn't really an instance of this?

There are more possible examples, but I don't want to strain the metaphor too much. In any case, the reproduction strategies and lifecycles of mememes are varied and often not comparable to infectious organisms at all.

One interesting parallel with biological reproduction and memetic reproduction mechanisms is lateral gene transmission. Bacteria (among others) can share genetic material "laterally", not just asexually but between two living members of different species. Mememes, and especially complexes of mememes ("memeplexes") can integrate content from others, rather than by direct inheritance, eg. Doc Martens

boots being created as a work boot and adopted as a punk fashion symbol, or traditional western militaries integrating strategies from insurgency movements. Besides lateral transmission, memes have other interesting kinds of reproduction and evolution, eg. memeplexes can merge with others, say in syncretic religious traditions.

Memes live in, or run on, human minds. They feed off of mental and emotional energy, and they compete with other memes for attention, or "mindspace."

Many memes can coexist, both in a single human mind and in a larger culture, and often exist in a kind of ecology. A "memetic monoculture" would die very quickly, a human culture can't consist *singularly* of weaving baskets, people need a reasonably large collection of practices, skills, relationships, and institutions to survive.

Below are some important kinds of memetic competition. Memes will starve or die out, if:

- they can't feed, meaning that their individual hosts stop practicing them or remembering them
  - eg., traditional crafts dying out, or secularism eroding religion
- their lineal transmission is stifled, hampered, or forbidden
  - eg., children no longer wanting to learn a minority language, or a practice becoming illegal or taboo
- they explicitly conflict with other memes
  - though, this can go different ways, and sometimes conflict can be energizing and sustaining

## Egregores

The term **egregore**<sup>1</sup> is used to refer to super-organisms or collective entities formed of many people. Egregores form when people see themselves and their community (understood broadly) as a cohesive identity, capable of collective action and needing protection. These can include religions, ethnic groups, corporations, social movements, etc. Egregores can often overlap with others, or might be contained completely within another larger one.

### **Egregores can be seen as eusocial memetic superorganisms**

The metaphor here goes as follows: multicellular organisms contain separate cells with their own metabolic structures cooperating to form tissues, then organs organs, and a larger organism. In particular, most individual cells are not gametes, and individually their direct lineage will end when the entire organism dies, regardless of if the organism successfully reproduced. Then, eusocial species (like bees) form large cooperative colonies where most individual members (workers) do not reproduce, even if the colony is successful and propagates.

Then extended further, egregores are like eusocial colonies *of memes*, not composed exactly of the human hosts, but composed of memes and existing in the collective mind-space of humans. Similarly to cells and worker bees, the memes will shift around, die, and be replaced, even while the larger super-organism (the egregore) keeps going, surviving both all of the individual human hosts and the memetic cells.

cell::multicellular organism::eusocial individual:eusocial colony::meme:egregore

Egregores are mostly "immortal" (in the biological sense, technically "non-senescent"), and like individual memes they can reproduce "laterally," as well as merging with other egregores. They do sometimes experience mitosis, but they don't usually reproduce sexually.

Egregores having something of their own identity, and "want" to persist, often despite the consequences for their host members, eg. stereotypical harmful cults. Egregores can slough off both component memes, and sometimes hosts. Sometimes these are flushed out like dead cells, but egregores can actually use controlled cellular death to *strengthen* the larger organism—almost every religion has its martyrs.

In an extreme form of this, egregores can often "outlive" whatever nominal organizing principles they represented at the time of their founding. Very common and powerful egregores in modern times are those constructed in terms of some ideology or religion, but often those egregores can persist despite an almost complete change in the contents of the ideology, whether by hostile takeover or by drift. An egregore can also sometimes persist through a complete turnover of its (human) members.

Nonetheless, egregores are still usually at least moderately cooperative with their hosts, and many are quite positive. In any case the basic pattern seems to be necessary for human survival and reproduction, unlike mostly solitary species like bears or spiders.

It's important to note that egregores don't need a particularly "coherent" ideology to function, and likely in the ancestral environment, and still in some parts of the world, egregores were structured mostly around kinship and diffuse communal recognition, without need for an ideology to justify them. A lot of the content and structure of ideologies has more to do with sustaining an egregore than being particularly adaptive beyond the functioning of the egregore, let alone "right" or "true." A particular shape of society often has some kind of homeostatic property to persist in that shape, among other things because on average it has been selected for to have such properties, and those that didn't have died out rapidly.

## For better and for worse

Broadly speaking, both memes and egregores are selected for (as in, differentially survive and reproduce) having properties that do not need to be closely aligned with our values personally or communally. Smallpox could persist for millenia despite being an obvious drag on productivity and wellbeing, but remained endemic until the development of modern vaccinationology.

More specifically, because memes and egregores "live in" or "run on" human minds, they can leverage all sorts of patterns of belief and feeling to justify (and protect) themselves. These justifications neither need to be aligned with the effects of those memes on their hosts, nor again even "true" or coherent. In particular, often the contents of a [metaphysical system](#), or individual metaphysical claims, are part of the buttresses and defense mechanisms of memes and egregores.

**The fitness or persistence of a meme does not necessarily have much to do with its wholesomeness or helpfulness for its human hosts.**

A meme, memplex, or egregore can't be *too* toxic for its hosts, lest it die off with them, but like many endemic infections it can be a major burden on the health and wellbeing of the host species while still allowing them to carry on. Sometimes this kind of burden can be seen as a kind of "excess" or "miscalibration", but often the toxic effects of infectious agents are part and parcel of their reproductive strategy (both for biological organisms and memetic ones).

cursedness

memetic justice? autonomy?

## Footnotes

1. idk where this stuff came from, 🤖 it's in the zeitgeist. ↔ ↔<sup>2</sup>
2. IIRC technically not organisms. ↔

# Culture, Taboo, and Trauma

---

trauma, taboo, and cultural trauma



# Burdens of Meaning

---

# Discourse

---

# Spirituality, Phenomenology, and Practice

---

*In this section:*

# What are Spirituality & Spiritual Practice?

---

- something like, this is opinionated and sort of just my schtick. I'll try to be narrowly descriptive but one way or another that you cut this up ends up prioritizing different aspects. Trying to achieve some like "cross-anthropological" perspective is presumptive and in any case just another view
  - again, nonetheless, I try to be descriptive and accurate as far as it goes
- my opinionated, hypermodern take here is that devotional and mystical stuff is a. largely contiguous with ordinary experience, b. would have been more common in times past, and c. is not really distinguishable from a bunch of other stuff, eg talmudic study is part of religion but what about ordinary legal study, mathematical study? lots of things arouse awe and devotion: group dance/chanting re sports events, esp music at sports events, music in general, superhero movies, etc etc.
  - and then it's just like, yeah, there's a bunch of these emotional and psychological states, they can be more altered, they can be more intense, they can be more spacious, submissive, devotional, etc; and in the usual modern view they're taken as mostly a separate domain but I don't buy this
- ghost touched my honker story (not the ghost but seeing god etc); devotional practice; mystical attainment: "held int he hands of the buddha/god" mooji peace pilgrim, harada roshi

# Everybody's Into Transformative Practice

---

# Global Wayfinding

---

*In this section:*

# Layering and Technical Debt

---

[Global Wayfinding](#) is a model of meditation and transformative practice, developed by Mark Lippmann. Mark's materials are unfortunately fairly inaccessible, often even to people with both substantial technical *and* meditation backgrounds. I'd like to make some concepts more available which haven't yet made their way into the discourses I'm in.

Claims in this essay are meant to accurately reflect Mark's models, and I can't quite sharply endorse all of them. In particular I can't attest to many of these claims from my own practice, which is generally the epistemic standard I hold for claims about meditation/phenomenology etc. Also, my pool of anecdotal data is much smaller than Mark's, and I've forgotten a lot of the details of where various claims and models came from.

## Layering

In the Wayfinding framework, a person is said to *layer*, sometimes *layer over*, when a compensatory structure is added to prevent, diminish, or attenuate perceived negative parts, behaviors, or emotions. Layering can be variously effective, sometimes completely blocking out a part or a class of parts from coming to conscious awareness. Layering is obviously *not* causing the structures responsible for the undesired qualities or states to cease, only for their arising in consciousness to be prevented. Sometimes layering is initiated "consciously" but often the layering that people report discovering in their practice is not willful, and they usually have little control over it consciously. Extant layering may also be quite old, and a meditator may have no available memory of the events which led to it. *Delaying* is a consistently noted phenomenon among serious practitioners. There are many reports of people "discovering" in the course of their practice intense sadness, anger, sexual energy, or many other qualities which might have been covered over or dissociated from.

Many traditional religious and meditative systems prescribe practices to purposefully attenuate various undesirable mindstates. These are variously effective for different people, and though they sometimes work to substantially attenuate negative mental qualities *in the moment*, they are generally not effective at ultimately dissolving those parts. I find it telling that teachers often prescribe practices to persistently push against negative states. Sometimes traditional teachers recognize that this reflects the fact that purposeful attenuation is incomplete, but it's claimed that after sufficiently deep awakening these negative states fall away of their own accord. I would call all of these kinds of practices "purposeful layering". Some kinds of layering practices are relatively gentle but still ultimately problematic, and many are quite forceful. Again I find it conspicuous that traditional teachers prescribe forceful layering practices for long periods of time. A similar critique applies to many contemporary schools of psychotherapy.

It should be said that layering is necessarily functional, and that even "correct practice" according to Wayfinding doesn't magically destroy all layering at once. Existing layering will persist until it is carefully unraveled, and new layering may accrue while the mind does not yet have better tools to accommodate or integrate negative experiences.

## Technical debt

*Technical debt* is a term from software engineering that refers, among some other things, to the accumulation of unmanaged complexity in a piece of software. A good image is that of the Burrow (the Weasleys' house) in Harry Potter. As new features are added, they are tacked on without attention to the structural integrity of the program. Each new feature is crammed inside, under, on top of, or dangling off existing structures. As this continues the piece of software becomes more and more difficult to maintain: imagine trying to fix the electricity in one room when it's got another room standing on stilts poking through its roof.

Technical debt relates to *refactoring*, another concept from programming, which is the process of tearing down a mess like this and creating a new architecture which is better suited to what's now needed.

Most people come to meditation practice with elaborate structures of layering. Mark's claim is that correct transformative practice consists of careful unraveling, of thorough refactoring of the person with perfect care for all of their values. In Mark's model, the mind is understood to be *lossless* with respect to its values (modulo brain damage, dementia, etc.), and so therefore "incorrect" attempts at dissolving negative structures will either simply be ineffective, or will result in a layered structure being tucked further away, thereby accruing more technical debt. Unlike a piece of software, however, a person cannot be safely shut down and taken apart, nor can one know ahead of time what the "correct" new architecture will be. Therefore, the process of Wayfinding is of an intricate back and forth dance, of healing and delayering, relayering, and allowing parts of the mind to come to peaceable solutions on their own terms, in their own time.

Lastly, according to Mark, layering is often quite subtle and intricate. My description of layering may evoke images of someone deliberately repressing anger, which is indeed a kind of layering. However, often a layering structure contains layers compensating for other layers, with small but pervasive effects left throughout the body and mind.



# Gnosis, Inference, Deference, & Confabulation

---

# Content, Healing, and Purification

---

# Practices and Traditions

---

# Schools, Teachers, and Transmission

---

# Spiritual Truth

---

*na hi verena verāni  
sammantīdha kudācanaṃ  
averena ca sammanti  
esa dhammo sanantano*

# Wrong Practice

---

My previous post on correctness typologies [ 2024-12-19 I'm not sure if that's what I'm gonna call it, and I might change the url] was meant to establish some distinctions around different kinds of "wrong" or "improper" ideas or actions. Whatever claims I make in this post are meant to be understood narrowly in the framework of correctness that I establish there.

I'll be explicit: I think there is such a thing as "wrong practice", and that much transformative and spiritual practice taught currently is wrong practice. First, however, I want to be clear about which kinds of "wrong" I think even apply in this domain, and which I mean in this case.

# Axiology

---

*In this section:*

# Harm

---



# Metaphysical Harms

---

- 2024-12-21 I want to discuss what harm is in principle, like, there's cost and there's harm. maybe this is its own post
- if you see your metaphysics as right, then deviations from it seem like weird disruptions of value
- rats will see this, hippies will see this. unclear if some secular normie would see this tho?
- this might end up going in the memetics section. anyway 2025-01-30 I might want to link back to rightness-ontology re misuse of rhetoric
- Lying is an epistemic harm, we understand that lying is bad! (Tho basically everyone does some amount of it) Similarly metaphysical harms, or at least one kind of metaphysical harm

# Metaphysics and Axiology

---

- moral claims
- goodness, not just "ethics"; "axiology" is sort of more helpful than "metaethics" imo (even tho they're often kinda the same)

**nihilism, nebulosity, and meaning**

# What is Axiology?

---