

Territory and the Maps

 putanumonit.com/2018/09/23/territory-and-the-maps/

View all posts by Jacob Falkovich

September 24, 2018

This is the first post of a sequence in which I lay out my philosophy of everything. Yes, you are very right to be worried. This is a terrible idea, and it tends to lead to confusion, madness, and worse – academese writing. And yet, the plan is for each of the following to eventually transform into a green link to a post:

- How I think about reality (this post).
- How I think about free will and desert.
- How I think about desires, motivations, and values.
- How I think about morality.
- How I think about happiness.
- How I think about meaning.

Already, you may be encouraged. Each section is titled “How *I* think about X” instead of “How *to* think about X” or, Aristotle forbid, “*What is* X”. My goal is neither to get an A in philosophy class nor to win an argument and own the libs.

This philosophy is meant to be useful in the edge cases of normal life. For quotidian things, one doesn’t need a philosophy at all, and on the other hand, I’m not particularly concerned with whether my philosophy works for hypothetical gods living in a simulation at the end of time. It is also the fundamental worldview that underpins my writing, and laying it out explicitly is an exercise in foundation-laying, [similar to this](#).

The main aim of my philosophy-of-everything is coherence, or *alignment*: alignment of my beliefs with reality, alignment of my actions with my values, alignment of my outcomes with my goals, and alignment of all the pieces of my philosophy with each other. This should really be its own post:

How I think about alignment.

This will all sound familiar to those who have read [Eliezer](#) and [Scott](#), and it may sound quite insane to those who haven’t. In either case, spilling endless pixels will not help. I will strive to be brief and to fill in the gaps with judicious links.

The first move in thinking about all of the above is understanding which aspects belong to reality, and which to the mind’s understanding of reality: the territory and the maps. Grasping this core distinction is enough to avoid the dumbest 50% of philosophical arguments, and yet it remains elusive for many people.

Reality is what determines the outcomes of experiments, the results of actions, the immediate experience of observing something. We call it ***the territory*** because it is a single whole. Minds contain representations of reality, or ***maps***. Maps are myriad,

representing the territory at different levels of abstraction, overlapping in some places and blank in others. You can map a person as a node in a social network, as an agent with goals, as a physical body with temperature and mass; those can all be useful in different contexts, but territory underlying them is one.

The first step in dealing with any concept is identifying whether we should think of it as a property of the territory, a property of maps, or a property of the process of *mapmaking*.

Truth, or accuracy, is a property of maps. Reality isn't "true", it just is. "Snow is white" is true if and only if snow is white. *Snow* is part of the territory – it can be white or not white. "*Snow is white*" is part of the map – it can be true or not true.

A map is accurate if it reflects the territory and allows you to traverse it – predict and act upon reality. This measure of truth folds together both epistemic rationality (good prediction) and instrumental (effective action). I do it for two reasons:

1. The conflict between truth and "winning" is vastly overstated. In practice, what's correct is what's useful.
2. According to predictive processing, most of what your brain does is accurately predicting sensory inputs in order to guide action. The two (sensory prediction and action) happen in a single, continuous process.

Predictive processing, which I'm a fan of as a model of cognition, offers its own take on reality: *reality is what generates the prediction errors*.

The mind doesn't live inside itself – it deals with the outside world. When you throw a basketball at the hoop your brain starts from predicting the flight of the ball, not from thinking about the muscles in your arm, and definitely not from thinking about itself. The input your brain is most interested in is what the ball actually, which is the part that takes place in reality.

Indeed, our brains are so built to work with external reality that it takes a dozen years of schooling before anyone will think to ask whether objective reality is a thing. But this also raises an issue that confuses some people: properties of a mind appear inside that mind as properties of the territory.

"Jacob believes that snow is white" appears in my own head simply as snow being white (in the territory). But when *I* write that bananas are yellow, *you* interpret it as a fact about *my map*. I could have beliefs about my own map which are separate, such as my own belief that "Jacob believes that bananas are yellow".

One can have the first-order belief without knowing that you have it (although that makes it hard to talk about). Example: you believe that 101 is prime, but you didn't know you believed that until you checked. Whether it takes you 0.2 seconds or 20 minutes, your map will almost certainly represent "101 is prime" when queried, the same way a street exists in Google maps even if it's not currently on your screen.

Alarmingly, you can think that you believe something when, in a very real sense, you don't. This is called **belief in belief**.

Don't worry about being lost in the meta-layers. We will mostly stick to reality and the two levels above it: the territory, a map, and a map of a map.

Let's take a quick tour of *stuff* and place everything where it belongs in the scheme.

First, the **unconscious mind**: one's intuitions, urges, and perceptions. My philosophy-of-everything is a mapmaking exercise that sits squarely in the conscious and analytic system 2, and so the systems 1 are **part of the territory** as far as I'm concerned. I can make observations of what my unconscious does, but I do not rewrite it directly as I would a conscious map.

Rationality is a property of **mapmaking**, not of the map itself. A method of reasoning and evaluating evidence is rational if it tends to arrive at accurate pictures of reality. Having a true belief isn't *proof* of a rational process even though, according to the rational algorithm of Bayes' theorem, it's *evidence* of such a process.

If you think that the hot hand effect in basketball is real because you carefully analyzed the data you won't be any more correct than someone who thinks the same way because they heard a sports announcer say it. But you're much more likely to be correct about other things in the future by ignoring sportscasters and being careful with statistics.

Reality has patterns, but **categories** are only borders drawn on maps. Whether a tree falling in the forest makes "a sound" depends on the purpose of the map. So does whether a whale is a fish – are you doing phylogenetics or sailing? This doesn't mean that all categorizations are equally valid. If you say that a whale is a tree when it fishes soundlessly, I'll ask Wittgenstein to punch you.

Some categories are so central to how our brain works that it's hard to distinguish them from reality. An important example of this is your selfhood, the sense of yourself as a distinct entity. It is clear that your "self" doesn't exist on the level of quarks and atoms, but there are many more places where you will find it missing.

A similar confusion happens with **purposes**. It is often very useful to treat something as if it is an agent with intentions, but we often go way way overboard with this. Little children intuitively assign purposes even to inanimate objects: the sun shines *so that* the trees would grow, the mountain is there *for climbing*. Some adults don't grow out of this. I recently had a discussion about gun control with a woman who rejected the potential relevance of statistics such as the number of homicides committed by non-gun objects and the number of non-homicide uses of guns because *the purpose of guns is murder*. She explicitly stated that the *murderness* is a property of the object itself, independent of what humans do.

It's much more common for people to mistake properties of the mind for properties of reality, but I think that the opposite happens with **values**. People seem to place consciously endorsed values on a pedestal, distinct from mere *desires*, *urges*, and *tendencies*. And yet if the **territory** contains an alcoholic who hasn't been sober in a year, their protestations that they "value sobriety" point to a weird definition of "value". More on that on the post about values.

I will use the map/territory/mapmaking in all the subsequent posts, but this shall do for an introduction. Stay tuned!