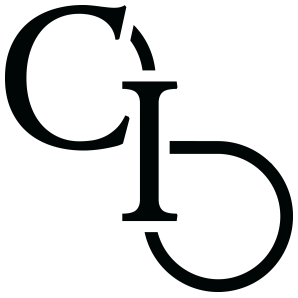
People, Reason, & Reality

Part I: Reason · Module 3

Explanations: Good, Bad, & Constrained

Logan Chipkin



CONJECTURE UNIVERSITY

Explanations: Good, Bad, & Constrained

Before the discovery of science and reason, earthquakes must have been both terrifying and mystifying. Who or what caused these unpredictable, destructive events? Why did the otherwise immovable ground occasionally shake so hard that it split apart beneath their feet?

The Greeks posited that earthquakes were caused by the god Poseidon's rageful strikes against the ground by his trident. This divine theory of earthquakes is undoubtedly *an* explanation of earthquakes. But is it a good one?

As philosopher and physicist David Deutsch explains in *The Beginning of Infinity* and elsewhere, a good explanation is one that is hard to vary while retaining its ability to explain what it purports to explain. An explanation typically consists of many components. If an explanation's explanatory power (or lack thereof) remains unchanged when the components are tampered with, then it is easy to vary and, therefore, bad.

Consider: why are earthquakes caused by Poseidon's bouts of rage rather than Zeus's? Why is it the strike of a trident that causes the tumult, and not any other weapon? Why does Poseidon strike the ground when he is angry, rather than when he is delighted? Swapping out the god's identity, the emotion that motivates the act of striking the ground, the weapon that hits the Earth, and a myriad of other components does not improve or spoil the original explanation; any such explanation 'works' just as well as the original.

This arbitrariness implies that there is no rational way by which to choose one set of components over any other. That is, there is no rational standard of criticism by which we might decide to adopt the 'angry Poseidon strikes ground with trident' theory of earthquakes over rival explanations such as the 'merry Zeus strikes ground with Master Bolt' theory.

In the seventh century BCE, pre-Socratic philosopher Thales of Miletus (whom we will meet again when we discuss the significance of a *tradition of criticism*) offered an entirely naturalistic explanation of earthquakes. He posited that the (flat) Earth floated in an infinite ocean, and that the occasional violent water waves caused the Earth to slosh around and suffer earthquakes. While this explanation had the advantage of being

impersonal as compared with the Poseidon theory of earthquakes, it was still not a good explanation—Thales' explanation was also easy to vary. For example: why should the Earth float in a body of *water*, rather than any other fluid?

Today, we explain earthquakes via the theory of plate tectonics: massive, rocky plates beneath the Earth's surface sometimes press up against one another along fault lines, building up large amounts of stress. Most of the time, friction between the plates prevents them from slipping, but sometimes a plate slips. When this happens, the stress energy is released in the form of seismic waves, resulting in earthquakes. Is this explanation easy to vary, like its predecessors? Look at the explicit components in this explanation: rocky plates, fault lines, stress, energy, friction, seismic waves. Could I have replaced any of them without spoiling the explanation? Whereas Poseidon could have been replaced with Zeus, and Thales' water could have been replaced with milk, the rocky plates *cannot* be arbitrarily replaced without the entire tectonic plates explanation of earthquakes losing coherence. Had the plates been made of cheese, for example, then they could not build up the crucial stress energy that I'd described earlier. Nor can we arbitrarily *remove* any component of the explanation. Had the plates not been separated by fault lines at all, then there could be no slippage. Had there not been slippage, then there could be no release of energy. The plate tectonic theory of earthquakes consists of interdependent components, none of which can be arbitrarily replaced, none of which can be removed without the entire explanatory edifice crumbling into nonsense.

But could the plates have been made of *metal*, rather than rock, and left the explanation intact? Is rock really the only substance that coheres with the other aspects of the tectonic plate explanation of earthquakes? If the plates were made of metal, would they still be able to collide, build up stress, slip, and release seismic waves? We don't need to dive into material science here. For the sake of argument, let us assume that, yes, metallic plates could 'do the job' as well as rocky plates. Doesn't that refute my claim that the entire theory is hard to vary?

It would, if not for the fact that theories are constrained by all of the rest of our good explanations about the world. In this case, we have an altogether independent and good explanation for why the plate tectonics consist of rocks, rather than any other substance that might well do the job required by the plate tectonic theory of earthquakes. Therefore, good explanations must be hard to vary not only in the sense that their

components are not infinitely arbitrary, but also in the sense that they cohere with the rest of our good explanations of reality.

What if I offered the Neo-Darwinian theory of evolution as an explanation of earthquakes? It is definitely hard to vary: its fundamental concepts, such as genes, natural selection, variation, mutation, form a delicate yet coherent explanatory whole (Swap genes for species, natural selection for random drift, mutation for phenotypic change, and the Neo-Darwinian theory no longer makes sense.) It is also consistent with the rest of our good explanations across physics, biology, and biochemistry. And yet it *is* a bad explanation of earthquakes for the intuitive reason that it fails to explain them. An explanation can be hard to vary, cohere with the rest of our good explanations, and yet fail to explain what we want to explain.

An apparent explanation can also be consistent with all of our other knowledge but still easy to vary, thereby making it bad. For example, one may say, "Earthquakes are caused by natural events that precede earthquakes in time." That is true enough, but you can swap 'earthquakes' with any other physical phenomenon without spoiling this 'explanation'. Note that this explanation, while easy to vary, *would* be an improvement over *supernatural*, easy to vary explanations, such as the Poseidon explanation. But the absence of any particular error in an explanation—in this case, the philosophical assumption that regularities are explicable in purely naturalistic terms—is not enough for an explanation to be hard to vary. Moreover, as we see in this example, even being true is not enough.

To summarize, a good explanation is not only hard to vary, but it must also cohere with the rest of our explanations and actually explain what we are trying to explain (the technical term for the phenomenon we wish to explain is *explicanda*).

	Is this explanation hard to vary or easy to vary?	*Is this theory consistent with the rest of our (contemporary) good explanations?*	*Does this theory purport to explain earthquakes?*	*Is this theory a good explanation of earthquakes?*
Angry Poseidon theory	Easy to vary	Not Consistent	Yes	No
Thales' water theory	Easy to vary	Not Consistent	Yes	No
Theory of plate tectonics	Hard to vary (in light of our other knowledge)	Consistent	Yes	Yes
Neo-Darwinian theory	Hard to vary	Consistent	No	No
Theory of natural causes	Easy to vary	Consistent	Yes	No

These three constraints imply that the search for good explanations will always be nontrivial. In fact, the deeper our explanations of the world, the *more* constrained the space of good explanations becomes (we will revisit the concept of *constraints* in a future module).

But despite the inherent challenges in discovering a good explanation, we occasionally conjecture not one but two or more good explanations for the same explicanda. How can we possibly decide between them?

The answer depends on the *kind* of explanations they are. For example, selecting between two scientific explanations can require a vastly different toolkit than selecting between two epistemological explanations.

In the next module, we will explore the differences between the various domains of knowledge and explain how we choose between rival good explanations within each of them.

Thanks to Conjecture Institute Cofounder David Kedmey for valuable feedback.