3.2: Inference to the Best Explanation and the Seven Explanatory Virtues

Explanations help us to understand why something happened, not simply convince us that something happened (see chapter 1, section 1.3). However, there is a common kind of inductive argument that takes the best explanation of why x occurred as an argument for the claim that x occurred. For example, suppose that your car window is broken and your iPod (which you left visible in the front seat) is missing. The immediate inference you would probably make is that someone broke the window of your car and stole your iPod. What makes this a reasonable inference? What makes it a reasonable inference is that this explanation explains all the relevant facts (broken window, missing iPod) and does so better than any other competing explanation. In this case, it is perhaps possible that a stray baseball broke your window, but since (let us suppose) there is no baseball diamond close by, and people don’t play catch in the parking garage you are parked in, this seems unlikely. Moreover, the baseball scenario doesn’t explain why the iPod is gone. Of course, it could be that some inanimate object broke your window and then someone saw the iPod and took it. Or perhaps a dog jumped into the window that was broken by a stray baseball and ate your iPod. These are all possibilities, but they are remote and thus much less likely explanations of the facts at hand. The much better explanation is that a thief both broke the window and took the iPod. This explanation explains all the relevant facts in a simple way (i.e., it was the thief responsible for both things) and this kind of thing is (unfortunately) not uncommon—it happens to other people at other times and places. The baseball-dog scenario is not as plausible because it doesn’t happen in contexts like this one (i.e., in a parking garage) nearly as often and it is not as simple (i.e., we need to posit two different events that are unconnected to each other—stray baseball, stray dog—rather than just one—the thief). Inference to the best explanation is a form of inductive argument whose premises are a set of observed facts, a hypothesis that explains those observed facts, and a comparison of competing explanations, and whose conclusion is that the hypothesis is true. The example we’ve just been discussing is an inference to the best explanation. Here is its form:

1. Observed facts: Your car window is broken and your iPod is gone.
2. Explanation: The hypothesis that a thief broke the window and stole your iPod provides a reasonable explanation of
Notice that this is an inductive argument because the premises could all be true and yet the conclusion false. Just because something is reasonable, doesn't mean it is true. After all, sometimes things happen in the world that defy our reason. So perhaps the baseball-dog hypothesis was actually true. In that case, the premises of the argument would still be true (after all, the thief hypothesis is still more reasonable than the baseball-dog hypothesis) and yet the conclusion would be false. But the fact that the argument is not a deductive argument isn't a defect of the argument, because inference to the best explanation arguments are not intended to be deductive arguments, but inductive arguments. As we saw in chapter 1, inductive arguments can be strong even if the premises don't entail the conclusion. That isn't a definition of what an inductive argument is! As we've seen, in order to make a strong inference to the best explanation, the favored explanation must be the best (or the most reasonable). But what makes an explanation reasonable? There are certain conditions that any good explanation must meet. The more of these conditions are met, the better the explanation. The first, and perhaps most obvious condition, is that the hypothesis proposed must actually explain all the observed facts. For example, if, in order to explain the facts that your car window was broken and your iPod was missing, someone were to say offer the hypothesis that a rock thrown up from a lawnmower broke the window of your car, then this hypothesis wouldn't account for all the facts because it wouldn't explain the disappearance of your iPod. It would lack the explanatory virtue of explaining all the observed facts. The baseball-dog hypothesis would explain all the observed facts, but it would lack certain other explanatory virtues, such as "power" and "simplicity." In the remainder of this section, I will list the seven explanatory virtues and then I will discuss each one in turn. The seven explanatory virtues are:

1. **Explanatoriness**: Explanations must explain all the observed facts.
2. **Depth**: Explanations should not raise more questions than they answer.
3. **Power**: Explanations should apply in a range of similar contexts, not just the current situation in which the explanation is being offered.
4. **Falsifiability**: Explanations should be falsifiable—it must be possible for there to be evidence that would show that the explanation is incorrect.
5. **Modesty**: Explanations should not claim any more than is needed to explain the observed facts. Any details in the explanation must relate to explaining one of the observed facts.
6. **Simplicity**: Explanations that posit fewer entities or processes are preferable to explanations that posit more entities or processes. All other things being equal, the simplest explanation is the best. This is sometimes referred to as "Ockham's razor" after William of Ockham (1287-1347), the medieval philosopher and logician.
7. **Conservativeness**: Explanations that force us to give up fewer well-established beliefs are better than explanations that force us to give up more well-established beliefs.

Suppose that when confronted with the observed facts of my car window being broken and my iPod missing, my colleague Jeff hypothesizes that my colleague, Paul Jurczak did it. However, given that I am friends with Paul, that Paul could easily buy an iPod if he wanted one, and that I know Paul to be the kind of person who has probably never stolen anything in his life (much less broken a car window), this explanation would raise many more questions than it answers. Why would Paul want to steal my iPod? Why would he break my car window to do so? Etc. This explanation raises as
many questions as it answers and thus it lacks the explanatory virtue of “depth.”

Consider now an explanation that lacks the explanatory virtue of “power.” A good example would be the stray baseball scenario which is supposed to explain, specifically, the breaking of the car window. Although it is possible that a stray baseball broke my car window, that explanation would not apply in a range of similar contexts since people don’t play baseball in or around parking garages. So not many windows broken in parking garages can be explained by stray baseballs. In contrast, many windows broken in parking garages can be explained by thieves. Thus, the thief explanation would be a more powerful explanation, whereas the stray baseball explanation would lack the explanatory virtue of power.

Falsifiability can be a confusing concept to grasp. How can anything having to do with being false be a virtue of an explanation? An example will illustrate why the possibility of being false is actually a necessary condition for any good empirical explanation. Consider the following explanation. My socks regularly disappear and then sometime reappear in various places in the house. Suppose I were to explain this fact as follows. There is an invisible sock gnome that lives in our house. He steals my socks and sometimes he brings them back and sometimes he doesn’t. This explanation sounds silly and absurd, but how would you show that it is false? It seems that the hypothesis of the sock gnome is designed such that it cannot be shown to be false—it cannot be falsified. The gnome is invisible, so you can never see it do its thing. Since there is no way to observe it, it seems you can never prove nor disprove the existence of the sock gnome. Thus, you can neither confirm nor disconfirm the hypothesis. But such a hypothesis is a defective hypothesis. Any empirical hypothesis (i.e., a hypothesis that is supposed to explain a set of observed facts) must at least be able to be shown false. The sock gnome hypothesis lacks this virtue—that is, it lacks the explanatory virtue of being falsifiable. In contrast, if I were to hypothesize that our dog, Violet, ate the sock, then this hypothesis is falsifiable. For example, I could perform surgery on Violet and see if I found remnants of a sock. If I didn’t, then I would have shown that the hypothesis is false. If I did, then I would thereby have confirmed the hypothesis. So the “dog ate the sock” hypothesis is falsifiable, and this is a good thing. The different between a true hypothesis and a false one is simply that the true hypothesis has not yet been shown to be false, whereas the false one has. Falsifiability requires only that it be possible to show that the hypothesis is false. If we look for evidence that would show that the hypothesis is false, but we won’t find that evidence, then we have confirmed that hypothesis. In contrast, an unfalsifiable hypothesis cannot be confirmed because we cannot specify any evidence that would show it was false, so we can’t try to look for such evidence (which is what a rigorous scientific methodology requires).

Suppose, to return to my broken window/missing iPod scenario, that my friend Chris hypothesized that a 24 year old Chinese man with a Tweety Bird tattoo on his left shoulder broke the window of my car and stole the iPod. This explanation would lack the explanatory virtue of “modesty.” The problem is that the hypothesis is far more specific than it needs to be in order to explain the relevant observed facts. The details in any explanation should be relevant to explaining the observed facts. However, there is no reason to include the details that the thief was 24 years old, Chinese, and had a Tweety Bird tattoo on his left shoulder. How do those details help us to understand why the observed facts occurred? They don’t. It would be just as explanatory to say, simply, that it was a thief rather than to include all those details about the thief, which don’t help us to understand or explain any of the observed facts.

The explanatory virtue of “simplicity” tells us that all other things being equal, the simplest explanation is the better explanation. More precisely, an explanation that posits fewer entities or processes in order to explain the observed facts is better than an explanation that posits more entities and processes to explain that same set of observed facts. Here is an example of an explanation that would lack the virtue of simplicity. Suppose that all three of our cars in our driveway
were broken into one night and that the next morning the passenger’s side rear windows of each car were broken out. If I were to hypothesize that three separate, unrelated thieves at three different times of the night broke into each of the cars, then this would be an explanation that lacks the virtue of simplicity. The far simpler explanation is that it was one thief (or one related group of thieves) that broke into the three cars at roughly the same time. In the domain of science, upholding simplicity is often a matter of not positing new entities or laws when we can explain the observed facts in terms of existing entities and laws. My earlier example of the sock gnome stealing the socks vs. our dog Violet taking the socks is a good example to illustrate this. Sock gnomes would be a new kind of entity that we don't have any independent reason to think exists, but our dog Violet clearly already exists and since the observed facts can be explained by Violet’s actions rather than that of a sock gnome, the Violet explanation possesses the explanatory virtue of simplicity, whereas the sock gnome explanation lacks the explanatory virtue of simplicity. However, sometimes science requires that we posit new kinds of entities or processes, as when Copernicus and Galileo suggested that the sun, rather than the earth, was at the center of the "solar system" in order to explain certain astronomical observations. In physics new entities are often posited in order to explain the observations that physicists make. For example, the elementary particle dubbed “the Higgs boson” was hypothesized by Peter Higgs (and others) in 1964 and was confirmed in 2012. Much earlier, in 1897, J.J. Thompson and his collaborators, drawing on the work of earlier German physicists, discovered the electron—one of the first elementary particles to be discovered. So there is nothing wrong with positing new laws or entities—that is how science progresses. Simplicity doesn’t say that one should never posit new entities; that would be absurd. Rather, it tells us that if the observed facts can be explained without having to posit new entities, then that explanation is preferable to an explanation that does posit new entities (all other things being equal). Of course, sometimes the observations cannot be explained without having to change the way we understand that world. This is when it is legitimate to posit new entities or scientific laws.

The last explanatory virtue—conservativeness—tells us that better explanations are ones that force us to give up fewer well-established beliefs. Like simplicity, conservativeness is an explanatory virtue only when we are considering two explanations that each explain all the observed facts, but where one conflicts with well-established beliefs and the other doesn’t. In such a case, the former explanation would lack the explanatory virtue of conservativeness, whereas the latter explanation would possess the virtue of conservativeness. Here is an example to illustrate the virtue of conservativeness. Suppose that there are some photographs that vaguely seem to indicate a furry, bipedal humanoid creature that does not look human. My friend Chris offers the following explanation: the creature in those photos is Bigfoot, or Sasquatch. In contrast, I maintain that the creature in the photos is a person in a Bigfoot suit. Given just this evidence (the blurry photos), Chris’s explanation lacks the virtue of conservativeness since his explanation requires the existence of Bigfoot, which is contrary to well-established beliefs that Bigfoot is merely folklore, not a real creature. In contrast, my explanation possesses the virtue of conservativeness since there is nothing about someone dressing up in a costume and being caught on camera (or even someone doing so to play a practical joke or to perpetuate a false belief in a certain population) that conflicts with well-established beliefs. My explanation doesn’t require the existence of Bigfoot, but just the existence of human beings dressed up to look like Bigfoot.

It should be stated that some of the examples I have given could illustrate more than one explanatory virtue. For example, the example of the invisible sock gnome hypothesis could illustrate either lack of falsifiability or lack of simplicity. In identifying which explanatory virtues a particular explanation may lack, what is important is that you give the correct reasoning for why the explanation lacks that particular virtue. For example, if you say that the explanation isn’t falsifiable, then you need to make sure you give the right explanation of why it isn’t falsifiable (i.e., that there is no evidence that could ever show that the hypothesis is false). In contrast, if the explanation lacks simplicity, you’d have to...
say that there is another explanation that can equally explain all the observed facts but that posits fewer entities or processes.

Exercise

Identify which explanatory virtues, if any, the following explanations lack and explain why it lacks that particular virtue. If there is a better explanation, suggest what it might be.

1. Bob explains the fact that he can’t remember what happened yesterday by saying that he must have been kidnapped by aliens, who performed surgery on him and then erased his memory of everything that happened the day before returning him to his house.
2. Mrs. Jones hears strange noises at night such as the creaking of the floor downstairs and rattling of windows. She explains these phenomena by hypothesizing that there is a 37-pound badger that inhabits the house and that emerges at night in search of Wheat Thins and Oreos.
3. Edward saw his friend Tom at the store in their hometown of Lincoln, Nebraska just an hour ago. Then, while watching the World Cup on television, he saw someone that looked just like Tom in the crowd at the game in Brazil. He hypothesizes that his friend Tom must have an identical twin that Tom has never told him about.
4. Edward’s friend Tom died two years ago. But just yesterday Tom saw someone who looked and spoke exactly like Tom. Edward hypothesizes that Tom must have come back to life.
5. Edward’s friend Tom died twenty years ago when Tom was just 18. But just yesterday Edward saw someone who looked and spoke exactly like Tom. Edward hypothesizes that Tom must have had a son that he did not know about and that this person must have been Tom’s son.
6. Elise has the uncanny feeling that although her family members look exactly the same, something just isn’t right about them. She hypothesizes that her family members have been replaced with imposters who look and act exactly like her real family members and that no one can prove that this happened.
7. John thinks that since something cannot come from nothing and since we know there was a Big Bang, an all-powerful but invisible and undetectable being must have been the cause of the Big Bang.
8. Erin feels that she is being followed. Every time she looks over her shoulder, she sees someone duck behind an object to avoid being seen. She hypothesizes that it must be her 5th grade teacher, Mr. Sanchez.
9. While walking through the forest at night, Claudia hears some rustling in the bushes. It is clear to her that it isn’t just the wind, because she can hear sticks cracking on the ground. She hypothesizes that it must be an escaped zoo animal.
10. While driving on the freeway, Bill sees the flashing lights of a cop car in his rear view mirror. He hypothesizes that the cops must have finally found out about the library book that he never returned when he was in fifth grade and are coming to get him.
11. While driving on the freeway, Bill sees the flashing lights of a cop car in his rear view mirror. He hypothesizes that the cops are going to pull someone over for speeding.
12. While driving on the freeway, Bill sees the flashing lights of a cop car in his rear view mirror. He hypothesizes that the cops are going to pull someone over for going 13.74 mph over the speed limit.
13. Stacy cannot figure out why the rat poison she is using is not killing the rats in her apartment. She hypothesizes that the rats must be a new breed of rats that are resistant to any kind of poison and that evolved in the environment of her apartment.
14. Stacy cannot figure out why the rat poison she is using is not killing the rats in her apartment. She hypothesizes that the rats must be a new breed of rats that are immortal and that evolved in the environment of her apartment.
15. Bob is fed up with his life. He intends to kill himself so he gets his gun, puts bullets into it and pull the trigger. Miraculously, he is not killed. Bob hypothesizes that he must be immortal.