BEWARE THE CONVINCING EXPLANATION

G. Randolph Mayes

It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

– Sherlock Holmes, *A Scandal in Bohemia*

Most advice for sharpening our thinking skills concerns how to avoid bad arguments. But argument is only one of the two basic forms of reasoning. The other is explanation, and it is equally susceptible to abuse. You may already be familiar with certain forms of explanatory malfeasance. One of the best known is circular explanation, in which the stated cause is just a different way of describing the effect. (Why did we lose the match? Because our opponents scored more goals.) Here I'd like to introduce you to a less appreciated error of explanation. To my knowledge it has no widely accepted name, but I call it 'the convincing explanation'.[1]

Explanations (as well as arguments) can be convincing in a general sense for any number of dubious reasons. Sometimes it is because they issue from an apparently authoritative source, even though we have no idea what is being said. Other times it is because they are pleasingly simple, making us feel like we understand something that we really don't. However, neither of these is what I mean by a convincing explanation. To show you what I do mean I first need to be clear about what an explanation is, and how it differs from an argument.

In formal logic the term 'argument' refers simply to a list of premises followed by a conclusion. All philosophers know this definition, but we only rarely use the term in this highly restricted way. Rather, we almost always think of an

doi:10.1017/S147717561100011X © The Royal Institute of Philosophy, 2011

https://doi.org/10.1017/S147717561100011X Published online by Cambridge University Press

argument in terms of its typical cognitive or epistemic function, which is to provide support for believing the conclusion to be true.

Explanations are like arguments in that they are sets of premises followed by a conclusion, but they have a completely different function. Explanations do not provide grounds for belief; rather they help us understand why something we already believe is so. For example, when you ask why your friend Toni is angry, you are not looking for a reason for believing she is angry. You know this already, but you don't *understand* it; you want to know why.

Carefully note what I mean by the term 'conclusion' here, as it is slightly unintuitive. The conclusion is not necessarily a statement being demonstrated or proven by the premises. That would tend to restrict the term to the context of argument alone. A conclusion is just a statement that follows from the premises. In the context of explanation, the conclusion is a statement we already accept as true. The explanatory premises support their conclusion by providing causes of (the corresponding) fact.

An easy way to remember the distinction between argument and explanation is to think of them as addressing fundamentally different questions. The typical argument question is: How do you know this? The typical explanation question is: Why is this so? How do you know Toni is angry? She is swearing and throwing things at Jake. Why is Toni angry? She discovered that Jake is seeing another woman.

If you follow what I have said so far, then you understand that an explanation, by definition, takes its conclusion for granted. Indeed, this is something that every competent speaker already knows at an intuitive level. If you were to approach a man on the street and ask him to explain why he stole your wallet, he would immediately detect that you are taking it for granted that he stole your wallet. (This, you may have noticed, is similar to the old question: Have you stopped beating your wife? Logicians call this a 'loaded' or 'complex' question, examples of which

https://doi.org/10.1017/S147717561100011X Published online by Cambridge University Press

commonly involve building an illegitimate assumption into an explanation request).

Because an explanation necessarily presupposes the truth of its conclusion, it follows that an explanation is not logically capable of *convincing* us of its conclusion. If you understand this, the following definition may now strike you as a logical absurdity:

> A convincing explanation is an explanation that succeeds in convincing us that its conclusion is true.

This is the definition I want you to bear in mind, however. For even though the notion of a convincing explanation is logically absurd, the fact is that explanations often do succeed in convincing us of their conclusions in the absence of any real evidence at all.

Some examples of convincing explanations

In his extraordinary new book *How We Reason*, Philip Johnson-Laird illustrates how our preference for explanatory relationships can cause us to become convinced of things that are not even logically possible. Suppose you know that Paolo went to get the car, and that this task should take about five minutes, yet ten minutes have elapsed and Paolo has not returned. What is more likely to have happened?

(1) Paolo had to drive out of town.
(2) Paolo ran into a system of one-way streets and had to drive out of town.

If you are like most people, you will choose the second option. But on further reflection it is easy to see that this can't be right, since the second option actually contains the first. In other words, (2) can not be more likely than (1), since (1) has to be true anytime (2) is true. (This is called the 'conjunction fallacy'). Why is this mistake so common? There isn't a universally accepted answer to this question,

https://doi.org/10.1017/S147717561100011X Published online by Cambridge University Press

but one striking feature of (2) is that it doesn't merely state that Paolo drove out of town; it explains why. It seems to make sense of the situation; we feel that we understand it better, and we become convinced that a statement is more likely than it is logically possible for it to be!

Consider a similar example.

(1) Crime is increasing nationwide.
(2) Crime is increasing nationwide due to widespread poverty brought on by the global recession.

Now, which of these do you find to be more convincing with respect to the claim that crime is increasing? You may find neither one of these very convincing, but most people would say that (2) is substantially more convincing than (1). After all (1) is simply a bald assertion, whereas (2) provides reasoning in support of the assertion in terms of a known fact, the global recession. What we fail to notice, however, is that the reasons provided in (2) are causes, not evidence. Evidence for the claim that violent crime has been rising has only one respectable form, and that is crime statistics from representative samples of the population. What we have been given here is not an argument, but a causal explanation which takes the increase in crime for granted. (As you may know, crime has actually been steadily decreasing in the U.S. for a few decades now.)

Another nice example of convincing explanation comes from economics. Suppose you hear someone claim that an increase in the minimum wage will be harmful to low-income workers. This might strike you as not obviously true, especially since the intent of such an increase would be to help low-income workers. In this case it would be natural to ask for some evidence. Suppose you were offered the following reasoning:

> Any increase in the minimum wage will raise the cost of doing business. Hence, to maintain their

https://doi.org/10.1017/S147717561100011X Published online by Cambridge University Press

> profits, employers simply lay people off and make the remaining employees work harder.

This view is widely taught in economics textbooks and it really makes tremendous intuitive sense. But again you may have noticed that no statistical evidence for the harmful effects of minimum wage laws was cited. Instead you have been given a convincing explanation for why these effects actually occur. But an explanation why they occur presupposes that they *do* occur, and that is the very point at issue. (The truth is that the evidence for the claim that minimum wage laws hurt low-income workers is far from convincing and much work done since the 1990's has contradicted it.)

You might reasonably suspect that I am overstating the case here. Granted, no statistical evidence is given, but doesn't the causal story provide at least some rational basis for believing the conclusion? This is actually a subtle matter, and I'll address it more completely below. For now, let me just suggest that whatever evidentiary significance the reasoning above may have, it is extremely weak. One way to appreciate this is to consider a different story, one in support of the conclusion that increases in minimum wage typically help everyone involved:

> Any increase in minimum wage will temporarily raise the cost of doing business. Employers typically respond to this by accepting a temporary reduction in profits and investing in the training and infrastructure needed to help their employees become more efficient and productive in the long run.

This also makes a great deal of intuitive sense, though it is equally defective as evidence. Especially when dealing with very complicated social phenomena, these 'just-so' explanations are quite easy to produce. The ones that fit nicely with our preconceptions about how the world works

https://doi.org/10.1017/S147717561100011X Published online by Cambridge University Press

cause us to become unreasonably confident in their conclusions.

Here is one last example.

> Our thinking skills are eroding at an alarming rate as a result of our constant immersion in superficial multi-tasking activities. Our habits of e mailing, texting, updating our status on Facebook and Twitter and mindless channel surfing all conspire to distract us from the careful and critical habits of mind required to grasp complicated issues and engage in meaningful social relationships.

This is a highly convincing explanation, partly because most of us will agree that we spend a great deal of time engaging in precisely these behaviors. Still, the causes it lists beg the question whether the phenomenon exists. Are our critical thinking skills eroding? If so, this should be reflected by poorer performance in an array of standardized tests, fewer qualified applicants for jobs, generally declining productivity, poorer quality of science, diminished creativity, unhappier relationships. Unequivocal evidence for any of these is actually very hard to find. (Nicholas Carr's new book (*The Shallows: What the Internet is Doing to our Brains*) has been widely criticized precisely for explaining facts that have not been established.)

Interestingly, the convincing power of explanation is not limited to contexts in which explanatory information is being supplied by someone else. It even occurs when we are simply making up the explanations ourselves. For example, it has been shown that when people are asked to explain why risk takers might make better firefighthers they end up being more convinced than before that risk takers do make better firefighters.

Once you have been sensitized to this mechanism you will start to see it in more places. For example, it is easy to observe it in advertisements designed to convince people of the effectiveness of health remedies. Many widely used

https://doi.org/10.1017/S147717561100011X Published online by Cambridge University Press

treatments (not just 'alternative' therapies, but conventional medicines and diagnostic procedures) have no demonstrated benefit beyond the placebo effect. Since the evidence is so unimpressive, it is rarely cited. Rather, testimonials and pictures explaining how the treatment actually works are provided instead. Another particularly troubling example of convincing explanation is the documented tendency of jurors to be more convinced by a satisfying explanatory story of how a crime could have been committed rather than dry forensic evidence that it actually was committed in that way.

A deeper account

As noted above, the thoughtful reader may harbor some doubts about my account of the convincing explanation. One question she might ask is this: Are explanations and arguments really that different? After all, a good explanation of the fact that my toe hurts is that I dropped a hammer on it. But the fact that a hammer is falling toward my toe seems an excellent reason for believing that my toe is about to hurt. Similarly, the explanation of Toni's tantrum may be her discovery that Jake is cheating on her. But wouldn't my discovery of Jake's cheating ways serve as a good argument that Toni will be angry when she finds out?

This is an important question. It doesn't challenge the idea that argument and explanation have different cognitive functions. But it suggests that the difference isn't practically important because good explanations can be good arguments as well. To deal with this question properly I need to say a little more about how explanation differs from argument.

In the old days it was widely held that a good explanation is one in which the causes provided should be sufficient to produce the effect. In fact, one of the main reasons for insisting on this was the belief that good

https://doi.org/10.1017/S147717561100011X Published online by Cambridge University Press

explanations will serve in both an argumentative and a predictive capacity. (This expectation is apparent in our own tendency to ridicule so-called experts who glibly explain things – like the global financial meltdown – even though they fail utterly to anticipate them.) So it seems like an alternative way of analyzing the examples I gave above is just to say that they would be bad explanations even if the conclusions were true.

There are a couple of serious problems with this view. The first problem is that explanations that provide causally sufficient conditions aren't the norm. It is widely accepted that complicated physical, biological and social phenomena are not susceptible to deterministic explanation. For example, a good explanation of the fact that someone is infected by HIV may be that he or she had unprotected sex with an infected individual. But the truth is that having unprotected sex with an HIV infected person only slightly raises the likelihood of infection. Similarly, the best explanation of the fact that you ran over a pedestrian may be that you were texting while driving, though this behavior does not make it likely that you will run over someone on any particular occasion. (That's why people keep doing it.)

The second problem is that even causally sufficient explanations can function as arguments only after evidence for the causal relationship has been established. This is a point that the great Scottish philosopher David Hume put quite a bit more eloquently as follows:

> *Let an object be presented to a man of ever so strong natural reason and abilities; if that object be entirely new to him, he will not be able, by the most accurate examination of its sensible qualities, to discover any of its causes or effects. Adam, though his rational faculties be supposed, at the very first, entirely perfect, could not have inferred from the fluidity and transparency of water that it would*

https://doi.org/10.1017/S147717561100011X Published online by Cambridge University Press

suffocate him, or from the light and warmth of fire that it would consume him.

In practical terms what this means is that a cause can only function as evidence for a particular claim if the causal relationship itself has been established by previous evidence. If there is in fact a genetic mutation that causes autism, the existence of this mutation will not constitute an argument that a child is autistic until the evidence for this causal relationship has actually been established.

So, in sum, the distinction between argument and explanation ultimately rests on the following facts.

(1) We have different criteria of adequacy for arguments and explanations. Very good explanations of accepted facts, meaning explanations that raise our understanding a great deal, typically fall far short of providing the evidence required to accept them as facts in the first place.
(2) Explanations serve as arguments only insofar as they are indicative of a previously established evidential relationship. While well-established explanations are sometimes capable of doing double duty as arguments, this is not their primary function, and routinely permitting explanations to perform this duty will result in serious mistakes in reasoning.

Sherlock Holmes attributed his ingenious inferences to the power of deduction. Deduction certainly played a role, but he was more often engaging in what we now call 'abduction', or 'inference to the best explanation'. The quotation at the beginning of this essay shows that Arthur Conan Doyle was keenly aware of the misleading power of the convincing explanation. Holmes' here cautions his friend Watson that the legitimate purpose of a theory is to explain well-

https://doi.org/10.1017/S147717561100011X Published online by Cambridge University Press

established facts, not to create them. It is a warning we all do well to heed.

Thanks to Jim Simpkins, Emrys Westacott, Mark Alfino, John Peloquin, Scott Merlino, Randall Marrett and Stephen Law for their helpful comments on earlier drafts of this essay.

G. Randolph Mayes is Associate Professor at the Department of Philosophy, University of Sacramento. mayesgr@saclink.csus.edu

Note

[1] Psychologists have identified a related phenomenon called the 'explanation effect' in which people become more convinced that a phenomenon does occur after making up an explanation of how it could occur.

https://doi.org/10.1017/S147717561100011X Published online by Cambridge University Press